UNIVERSITY OF
KWAZULU-NATAL

COMPUTER AIDED TECHNIQUES FOR THE ATTRIBUTION OF ATTIC BLACK-FIGURE VASE-PAINTINGS USING THE PRINCETON PAINTER AS A MODEL

by

Adrian John Ryan (Student Number 901331730)

Submitted in fullfilment of the requirements for the degree of Doctor of Philosophy in the Classics Programme, Faculty of Humanities, Development and Social Sciences.

January 2009

Promotors : Professor John Hilton
: Dr Gavin Hough

DECLARATION OF ORIGINALITY

I, Adrian John Ryan, do hereby declare that

i) The research reported in this thesis, except where otherwise indicated, is my original research.

ii) This thesis has not been submitted for any degree or examination at any other university.

iii) This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

iv) This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

   (a) their words have been rewritten but the general information attributed to them has been referenced;

   (b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

v) Where I have reproduced a journal publication of which I am the author, I have indicated in detail which part of the publication was actually written by myself alone and not by other authors, editors or others.

vi) This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed: ________________

Place : ________________

Date : ________________

TABLE OF MATHEMATICAL SYMBOLS

$\mathbf{a}, \mathbf{b}$ ::	vector	(1)
$\mathbf{x}$ ::	feature vector	(2)
$\mathbf{A}, \mathbf{B}$ ::	matrix	(3)
$\mathbf{A}^{-1}$ ::	matrix inverse	(4)
$\mathbf{A}^T$ or $\mathbf{x}^T$ ::	matrix or vector transpose	(5)
x, y ::	scalar variable	(6)
X, Y ::	random variable	(7)
$p(X = x)$ ::	probability that $X = x$	(8)
$p(x)$ ::	probability distribution for x	(9)
$p(x, y)$ ::	joint distribution over x and y	(10)
$p(x\|y)$ ::	conditional probability or distribution for x given y	(11)
$E(f(X))$ ::	expected value of a function f of a random variable X	(12)
$E_Y(f(X))$ ::	expected value of $f(X)$ over all possible values of Y	(13)
Σ ::	covariance matrix	(14)
$\hat{x}$ ::	arithmetic mean	(15)
$\hat{\mathbf{x}}$ ::	arithmetic mean vector	(16)
Γ, Δ ::	sets	(17)
$\Gamma \times \Delta$ ::	Cartesian product of sets	(18)

ABBREVIATIONS

ABV	::*Attic Black Figure Vase Painters* [Beazley, 1957].
ARV^2	::*Attic Red-figure Vase-Painters*[Beazley, 1968]
CVA	::*Corpus Vasorum Antiquorum*
Amazons	::*Amazons in Greek Art* [von Bothmer, 1957]
Vasenlisten	::*Vasenlisten zur griechischen Heldensage* [Brommer, 1973]
Para	::*Paralipomena* [Beazley, 1971]
Add^2	::*Beazley Addenda* [Carpenter, 1989]
LIMC	::*Lexicon Iconographicum Mythologiae Classicae* [Ackerman and Gisler, 1981-]

Acknowledgments

I would like to thank, first and foremost, my promotors, John Hilton and Gavin Hough for their patience, support and advice. In particular I would like to thank John Hilton for doing all in his power to provide relief from the department's busy schedule and thus allowing me the opportunity to complete this project, and Gavin Hough for long and detailed discussions and for allowing me access to the resources of Envirovision Systems. In addition I wish to thank Anne Mackay for spending many hours introducing me to the study of vase-painting and to the Princeton Painter, in particular. To Anne I also owe the debt of a large collection of photographs of the corpora of Exekias, Group E, the Princeton Painter and the Painter of Berlin 1686. I would also like to thank Olivier Commowick who made his intuitive LaTeXPhD thesis template available freely on the Internet.

The project was made possible through the financial assistance of a number institutions and people. First and foremost I wish to acknowledge the generous support of the Ursula Vogel Trust whose overseas bursary provided me with the opportunity to visit the Beazley archive and collect photographs from a number of museums in Europe. I would also like to thank the National Research Foundation for their financial support in the form of a PhD scholarship. Finally, I wish to thank my parents for much needed financial support, particularly where the funds from the scholarships did not meet the needs of the project.

ABSTRACT

Because of their abundance and because of the insight into the ancient world offered by the depictions on their decorated surfaces, Attic painted ceramics are an extremely valuable source of material evidence. Knowing the identities and personalities of the artists who painted them not only helps us understand the paintings, but also helps in the process of dating them and, in the case of sherds, reconstructing them. However, few of the artists signed their wares, and the identities of the artists have to be revealed through a close analysis of the style in a process called attribution. The vast majority of the attributions of archaic Attic vases are due to John Beazley whose monumental works set the stage for the dominance of attribution studies in the scholarship of Greek ceramics for most of the 20th century. However, the number of new scholars trained in this arcane art is dwindling as new avenues of archaeological research have gained ascendency. A computer-aided technique for attribution may preserve the benefits of the art while allowing new scholars to explore previously ignored areas of research. To this end, the present study provides a theoretical framework for computer-aided attribution, and using the corpus of the Princeton Painter - a painter active in the 6th century BCE - demonstrates the principal that, by employing pattern recognition techniques, computers may be trained to serve as an aid in the attribution process. Three different techniques are presented that are capable of distinguishing between paintings of the Princeton Painter and some of his contemporaries with reasonable accuracy. The first uses shape descriptors to distinguish between the methods employed by respective artists to render minor anatomical details. The second shows that the relative positions of cranial features of the male figures on black-figure paintings is an indicator of style and may also be used as part of the attribution process. Finally a novel technique is presented that can distinguish between pots constructed by different potters based on their shape profiles. This technique may offer valuable clues for attribution when artists are known to work mostly with a single potter.

PREFACE

This study aims to prove the concept that computers may be used as a tool in the attribution of Attic black-figure ceramics using techniques from machine learning and pattern recognition. The very nature of such a topic is multi-disciplinary. The author does not believe that a topic such as this could possibly be undertaken without due consideration being given to each of the disciplines from which it borrows. This necessarily means that the target audience is quite small. In part, this is born out of necessity since the nature of the problem requires considerable engagement with the methods and theories from pattern recognition, archaeology, and art history.

Understanding the computer-based methods, and the method in which the results are reported, requires basic knowledge of linear algebra, probability and statistics. Everything else has been derived from the basic principles. All the archaeological concepts used are defined in the introduction to chapter 1, and in the relevant sections of the subsequent chapters. While chapters 2,3,4 and 5 have considerable mathematical detail, the first sections of chapter 3,4 and 5 explain the art historical motivation and may be of interest to art historians, as will the appendix which provides a brief introduction to the art of the Princeton Painter. In addition, much of the material may be understandable to a more general audience by glossing the mathematical derivations without undermining the overall argument. In addition, each chapter has a summary that explains the most important findings of the chapter without using mathematical notation.

To facilitate reading of the dissertation under these circumstances, there are numerous cross-references which are formatted as follows:
chapter.section.subsection.subsubsection. The thesis comes with a CD that contains a pdf copy in which these cross-references are active hyperlinks to the relevant parts of the work. Getting the most out of this study is best achieved by reading the electronic version, or reading the hard-copy while using the electronic version to follow the relevant cross-references.

Throughout this dissertation, vases are referenced according to the museum in which they are housed, and where applicable, the Beazley publication in which they appear. Exceptions are made for Princeton Painter vases which are referenced according to museum and according to the catalogue at the end of the appendix (A.7) which lists the relevant details from *ABV*, *Para* and *Add*^{2}.

Contents

List of Figures

Chapter 1

Introduction

This study concerns the attribution of painted Attic ceramics (which for historical reasons, will be referred to as vases in this study) to the hands that painted them, and in particular proposes some computer-aided methods for doing so. This study will be based exclusively on the attributions either made by Sir John Beazley or those that he accepted as correct. It should be stated from the outset that this study does not aim to prove or disprove the attributions of Beazley, but merely that machines may be taught to attribute in the same way. Unlike the chapters that follow, most of this chapter does not require any highly specialised knowledge. Exceptions to this are section 1.3.2, particularly 1.3.2.2 which may require some archaeological background, and the literature review (1.4) assumes some familiarity with the subjects discussed under each of the headings. On the other hand, readers familiar with Classical archaeology can ignore the Background (1.1), although those without such a backround should read it as it explains much of the archaeological terminology used in the rest of the study

1.1 Background

Archaic and Classical Athenian[1] painted ceramics are amongst the most informative archaeological remains left by that society. In the first instance this is because, from the 7th century[2] on, Greeks started to depict narratives and iconography on their polychrome vases[3] on the surfaces of the pots. In the 6th century, particularly in Athens,[4] the technique of painting these narratives crystallised into a well-defined visual culture, the interpretation of which is the concern of most research on these objects. Understanding the visual culture can provide valuable insight into the society that produced these artefacts.

There are a number of reasons for this. One is that the scenes on the vases typically depict mythological stories. These can often supplement the literary record and occasionally even provide the only record of certain myths. For example, the scene of Jason being disgorged by a serpant, such as appears on a cup by Douris (figure 1.1)[5], is not attested in literature but appears on 5 vases and has been the subject of several papers including Kraeling

[1] The terms Athenian and Attic will be used interchangeably to describe black- and red-figure vases.

[2] In this dissertation, unless context suggests otherwise, all dates are BCE.

[3] The subjects of which were usually mythological.

[4] In other centres such as Corinth, the focus shifted from narrative to ornament.

[5] Vatican 16545; *ARV*^{2} 437.116 *Para* 375 *Add*^{2} 239; image (c) 1998-1999 Roy George(http://www.goddess-athena.org/Museum/Paintings/Argonauts/Jason_and_dragon_Douris_painter_f.htm.

[1971], Mackie [2001]. A second is that the scenes can sometimes shed light on the culture and politics of the time.[6] Third, Greek pots are abundant which means that they provide a lot of evidence for archaeologists allowing meaningful inferences to be drawn from the distribution of finds.

For these reasons, for Classicists to get the most out of these artefacts requires a sound understanding of the objects themselves and the visual culture that gave rise to the depictions on their surface. A problem that faced archaeologists at the dawn of the 20th century was that, while there appeared to be an artistic tradition amongst the painters of these wares, the artists themselves were anonymous - the vast majority did not sign their names on the vessels. Knowing the identity of these figures is desirable because it helps us better understand the development of the visual tradition, which in turn allows us to date the artefacts more precisely (this idea is explained in more detail in 1.3.1). John Beazley, the most important scholar in the attribution of vase paintings, is responsible for most of the attributions in 6th century Attic vases. Attribution of these artefacts to the artists who created them has been an important aspect of the study of vases, and dominated the field for a large part of the twentieth century. Before discussing attribution in detail, a brief discussion of some important terminology is provided.

1.1.1 Black- and Red- figure

Between the sixth and fifth centuries BCE, Athens dominated the trade of ceramics in the Greek world, and its wares are distributed widely in Greece and Italy, and Attic ceramics have been found as far as Egypt.[7] The larger vessels are rendered in one of two techniques: black- and red-figure. To explain the distinction between these two techniques requires a brief explanation of the way Attic polychrome vases were decorated. Three types of dilute clay are used as a paint to decorate the surface of the vase. This is done after the clay has dried but before it is fired. A dilute glaze is painted on surfaces that will turn black during firing, clay rich in iron is used to paint areas that will be dark red, and very fine primary clay is used to produce a white paint. Firing is in three phases. During the first oxidisation phase, the clay turns red. The kiln is then partially sealed and during this reduction phase, the surface of the vase turns black as the ferric oxide in the clay reacts

[6] For example changes in the iconography of Herakles in the 6th century BCE (Boardman [1972, 1975], Cook [1987a] and Boardman [1989]) and Theseus (Davie [1982, esp.pp.26-28] and Conner and Jackson [2000, pp.82ff].) in the late 6th century provide valuable supplementary evidence to the scant literary record concerning the imagery and propaganda of Peisistratos (tyrant of Athens 546-527) and his sons(527-510).

[7] for example at the Greek settlement at Tel Defenneh.

Figure 1.1: An Attic red-figure kylix by Douris from around 470 BCE showing Jason disgorged by a sea-dragon. This theme appears on 5 other vases but is unattested in literature.

Figure 1.2: A section of a black-figure type B belly amphora with the colours labled.

with carbon dioxide to form a dark ferrous oxide. When the kiln is again opened for the re-oxidisation phase, the areas that were painted with dilute glaze do not re-oxidise and remain black, while the surface of the vase that was not painted at all remains light reddish brown, a state referred to as 'reserved'. Thus, the painter has four colours at his disposal, the reserved light brown/red of the clay, the added red and white, and the black glaze. In addition, incisions into the clay are used as outlines.

Black figure was the dominant technique for most of the 6th century. In this technique, the background against which the scenes are depicted is reserved while the skin of the male figures is black, and that of women is white. Red and white are used for added details, such as hair, and to decorate the clothing and accessories (like armour). In addition, these added colours are used for subsidiary decorations like the lotus and palmette festoon (indicated in figure 1.2).[8] This dissertation is concerned almost exclusively with black-figure. An example of a black figure amphora is given in figure 1.2. Red-figure seems to be an invention of the late 6th century and becomes the dominant medium for Attic pottery in the fifth century and much of the fourth. The technique is almost a reversal of black-figure, whereby the background of the

[8]Detail of the reverse of Durban 1990.30 attributed by Cahn to the Princeton Painter.

scenes are black and the human figures are reserved. In addition, white, red and incisions are used to add variety to the decorations, although these are used more sparingly than in the earlier black-figure.

1.1.2 Vase Shapes used by the Princeton Painter

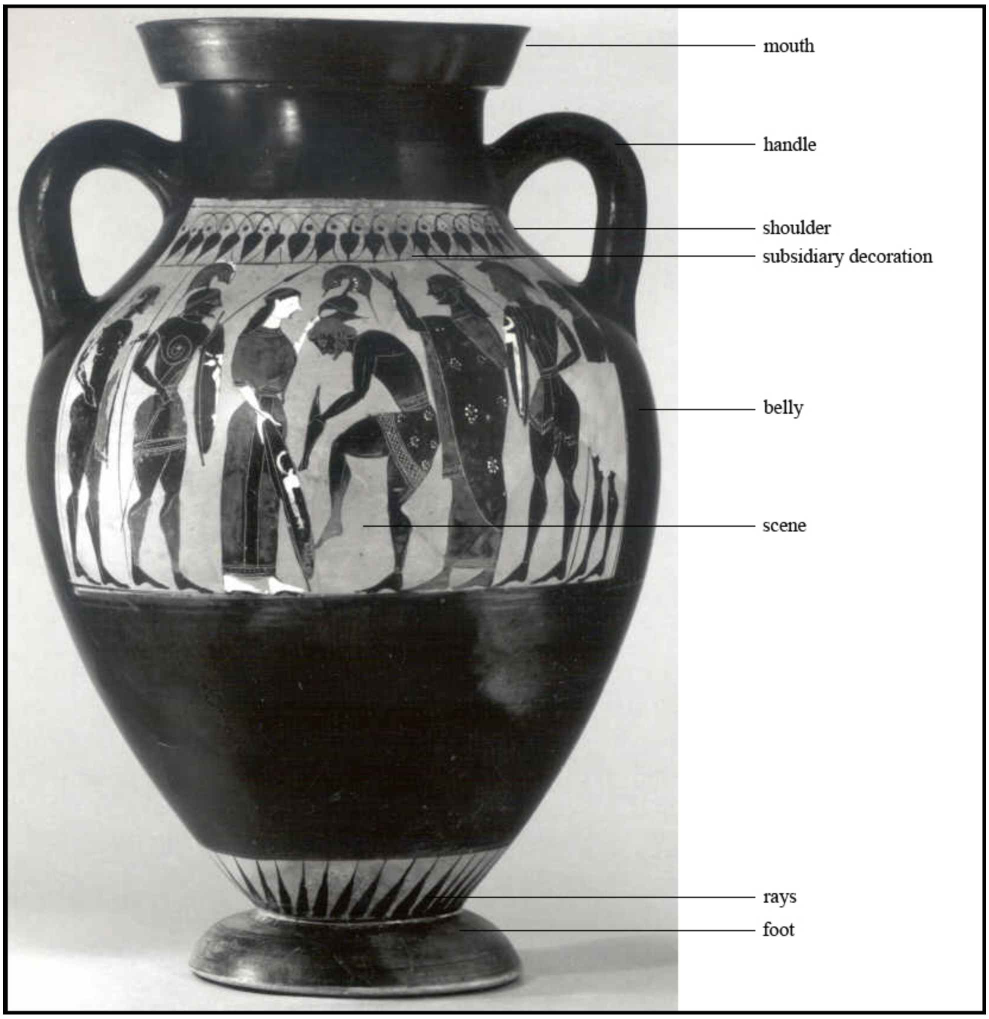


Figure 1.3: A typical 6th century belly amphora with parts labelled.

Figure 1.3[9] depicts a typical 6th century Attic vase with the different sections of the vase labelled. This may serve as a guide to this brief out-

[9]The example is Basel BS427 **EM1**, photograph courtesy of the Antikenmuseum Basel und Sammlung Ludwig.

line of the different shapes. Attic ceramics come in a variety of different types of shapes. Figure 1.4[10] illustrates those shapes that the Princeton painter employed. Briefly, the amphora was the most common shape for larger black-figure vessels.[11] Amphorae generally have a large belly and a narrower neck and vertically orientated handles that start on the shoulder or belly and terminate just below the mouth. Three distinct sub-shapes exist - the neck amphora which has the neck offset from the rest of the body, the belly amphora in which the body is a continuous curve, and the pelike which is similar to the belly amphora, but which has a very low centre of mass. The amphora is the most important shape for this study since the Princeton Painter, and indeed it would appear, most of the painters of the third quarter of the sixth centuries favoured this shape. In particular, the belly amphora is the most common extant shape in the corpora of all the painters discussed in this dissertation. In addition to the belly amphora, there is a special type of amphora called panathanaic. This shape is slightly narrower at both the neck and the foot than a standard neck amphora, and typically the point of articulation is less pronounced. These were usually filled with olive oil and given as prizes to victorious athletes at the Panathanaic games. Normally these have a standard scene on the reverse, with the legend *ton Athenethen athlon* (one of the prizes from Athens). The Princeton painter's variety are not of the true panathanaic type but instead copy the shape but without the standard decoration or inscription.

While the amphora is by far the most common shape in the Princeton Painter's extant corpus, there are some other shapes attributed to him by Beazley. The first is an oinochoe housed in the Antikenmuseum und Sammlung Ludwig in Basel. This vessel is a simple wine jug with one vertically oriented handle and usually a little spout for pouring. Beazley also attributed three hydriae to his hand. This is a vessel used to carry water. It has three handles, two of which are horizontally aligned and opposite each other, and one vertically aligned, which could be used to tip the vessel for pouring. In addition to these shapes, Mary B. Moore has attributed some fragments of a column krater to the Princeton Painter. The krater is a large two-handled vessel used for cooling wine. Unlike the amphora that has a small neck, the krater has a very large mouth, typically as large as the widest part of the belly. The column krater is a specific type in which the handles are vertically aligned.

[10]Figures C, D and E from Richter [1959, p.31], the rest by the author.

[11]In the Princeton Painter's corpus, neck amphorae are *ABV* 297.1-4; panathanaics5-6; belly amphorae are 7-21; hydriae are 22-4 and a lone oinochoe is 25.

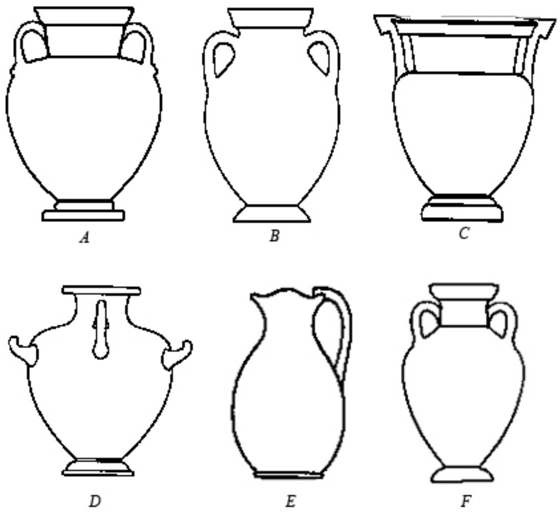


Figure 1.4: Outlines of vase-shapes used by the Princeton Painter. A) neck-amphora B)belly amphora C) column krater D) hydria E) oinochoe F) panathanaic amphora.

1.1.3 Attribution: Definition and Background

As has already been mentioned, attribution is an important part of the methodology of the study of Attic ceramics. Attribution is the process of identifying the author of a work of art, either because the work is unsigned or because there are doubts about its authenticity. For example, in the case of an unsigned painting, it is often important (for art historical as well as financial reasons) whether the painting is the work of, say, Vermeer, or simply one of his imitators. On the other hand, it is equally important to determine whether a signed painting is truly the work of, for example, Vermeer, or whether it is a forgery. Attribution is used for both purposes in the study of most Western art but in Attic black-figure is not used for authentication since other methods can be used for this purpose.

Attribution is a facet of the branch of art historical studies often called *connoisseurship*, which is limited to "identifying works of art with repect to date, provenience and authorship" [Panofsky, 1992, p.474]. The practice of attribution originated within the study of European art, and not archaeology. Until the late 19th century CE this was the province of learned authority figures who would make subjective pronouncements on the authenticity and authorship of works, which would usually be accepted on account of their personal standing. In 1880, this orthodoxy was challenged by Giovanni Morelli, a physician by training, who believed that a formal methodology would pave the way for a scientific approach to attribution.[12] Morelli believed that the key to attribution was a systematic examination of the way in which an artist rendered those minor details (which he called ***grundformen***) that did not require conscious reflection. Examples may be hands, eyes or minor drapery. He believed these were most likely the result of practice and therefore were less susceptible to the whims of the artists. Morelli advocated the compilation of catalogues of the diagnostic *grundformen* for different artists based on their known works, allowing the attribution of unsigned or spurious works to be conducted by comparing the *grundformen* on the subject with those in the catalogue.

Morelli's technique has subsequently been expanded by some of his successors. Berenson, for example, stressed that quality, a property difficult to catalogue in the same way as *grundformen*, should play a key role in attribution. Berenson advocated that the subjective intuition of the connoisseur was as important as the objective criteria, and that a combination of the two was needed for good attribution. Some connoisseurs like Max Friedlander, went further to say that attribution was actually achieved intuitively by the connoisseur who would then simply use the scientific method to justify their attributions [Friedländer, 1960, pp.166-7]. On the other extreme, van Dantzig [1973] advocated an empirical method called pictology by which a checklist of criteria is drawn up that completely describes an artists style. Attribution is then be achieved by assessing the degree to which the items on the checklist are satisfied.

While Morelli and his immediate successors had applied his methods mainly to Renaissance and later European paintings, connoisseurship first became a major tool in the analysis of archaeology in the early 20th century through the work of John Beazley. Beazley, who chiefly employed Morelli's method, would later be knighted for his monumental contribution of identifying the hands of over 1000 painters of Attic red and black-figure pots, classifying over 30000 vessels. While Morelli's technique is often used to detect forgeries, its use on Attic pots is exclusively for the purpose of identifying the hands of painters, since authentication of the vases is easy and may be achieved by other methods. Beazley's attributions were initially published in a series of articles and his master list was eventually published in monumental volumes (details follow in 1.1.4) that are now standard reference books for art historians and archaeologists studying Attic pottery. Beazley was

[12]Morelli first published (as Ivan Lermolief) a review of Italian works in Germany (Morelli [1880], first English edition is Morelli [1883]) and his method was expounded ten years later (Morelli [1890], first English edition is Morelli [1892]).

concerned almost exclusively with Attic pottery, so it was up to his students and followers to apply his techniques to the pottery of other centres, such as Corinth (Humphry Payne and Darrell Amyx) and South Italy (Arthur Trendall).

1.1.4 Conventions for Referencing Vases and Painters

Certain books and articles are referenced frequently in this study and in most studies of Attic vases. For this reason, they are discussed briefly here, as are the conventions by which these works are referenced in this thesis. In addition, a brief explanation is provided of the methods by which these painters are named and referenced.

While a handful of artists (such Exekias) have signed some of their wares and thus provide the Classical archaeologist with both a name and a base of examples from which to compile a description of his style, this is not the case for the vast majority of artists. Therefore, while scholars (particularly Beazley) may recognise that a group of vases are painted by the same person (or at least rendered in the same style), there is often no name associated with the respective artistic personality. Thus, names had to be invented for them. There are a number of conventional methods of naming these artists. Often this is done according to the museum in which the first of their pots to be attributed (called the **name vase**) is housed. For example, the Princeton Painter is named after a vase in the Princeton Art Museum (figure 1.5).[13] In some cases, the painter is actually named after both the museum and the catalogue number, as is the case with the Painter of Berlin 1686. Less commonly the painter is named after an unusual feature on one of his vases. For example the Swing Painter is named after a vase[14] depicting a girl on a swing.

The classification and establishment of taxonomies of artefacts and artefact types are central to archaeology and Beazley's spectacular success may be explained in part by his method's ability to impose order over the vast mass of undifferentiated pots and sherds residing in very diverse and disparate collections, and in part by the fact that Beazley did most of the work himself, thus providing the archaeological community with a ready reference to most of these artefacts. Beazley's lists are now standard reference works for the disciplines involved in the study of Attic vases, and furthermore his method so pervasive that few scholars publish substantial pieces of painted Attic pottery without some attempt to attribute them.

[13]The image is courtesy of the Princeton University Art Museum.
[14]Boston 98.918, *ABV* 306.41.

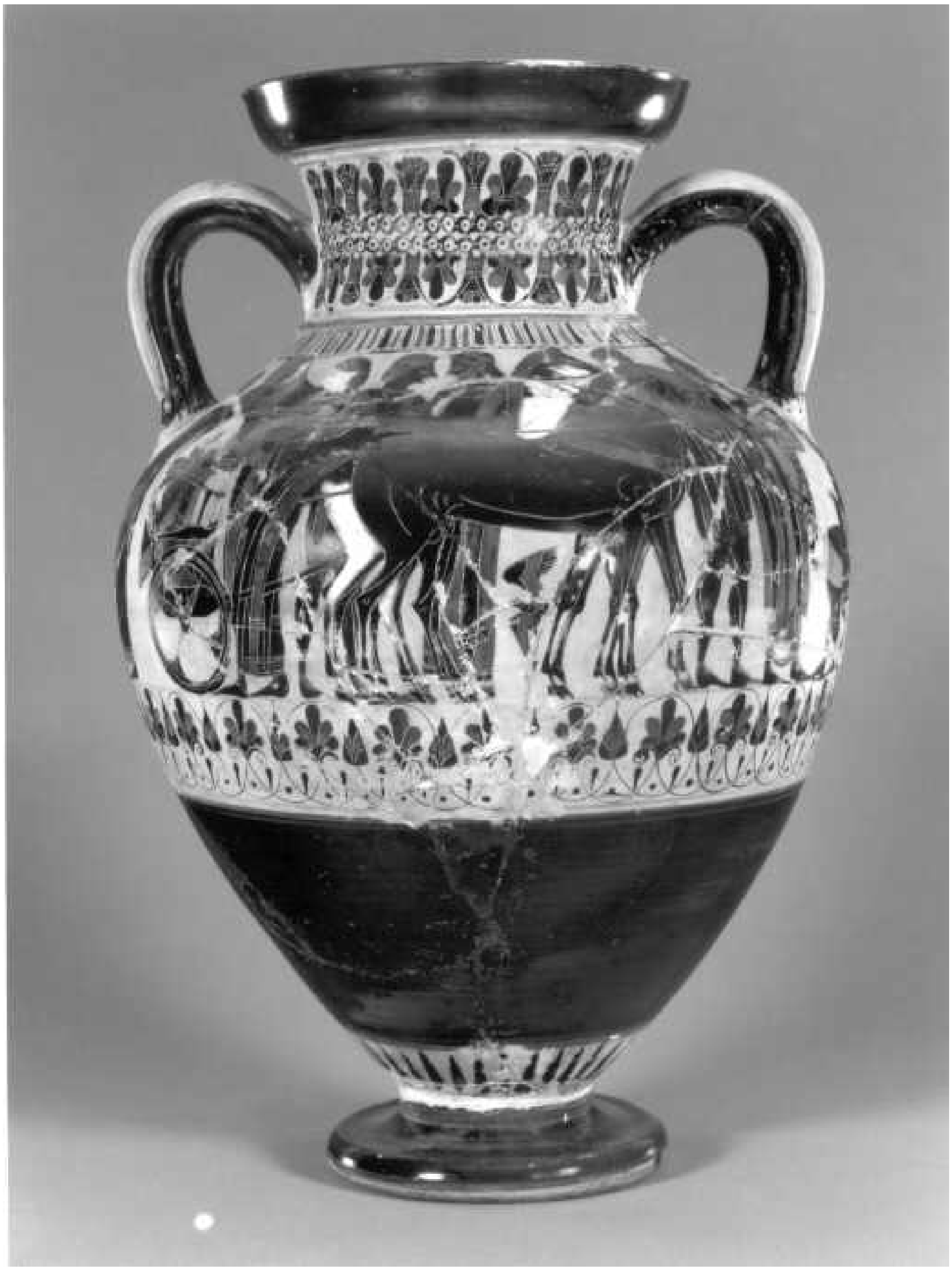


Figure 1.5: Princeton Y169, the name vase of the Princeton Painter.

Beazley's master list is the primary reference for most attributed Attic vases. This list contains the names of each painter he identified (and some identified by other scholars with whose attributions he agreed) followed by a list of the vases attributed to him, numbered according to shape. For red-figure, the bulk of his list is in the second edition of *Attic Red-figure Vase-painters*[15] abbreviated ARV^2 and for black-figure it is *Attic Black-figure Vase-paintings* (ABV). Both these books contain his lists as they stood at the time of publication, together with a bibliography of published photographs of each vase. Both these texts were later supplemented by *Paralipomena*[16] (*Para*) and *Addenda*, now in its second edition (Add^2)[17] each of which contain both updated bibliographies on the vases in ARV^2 and ABV and new attributions. *Para* and Add^2 keep the same vase numbers as ABV and ARV^2 but the page numbering is different. New additions to Beazley's list (i.e. those not in ARV^2 or ABV) are usually not given a unique new number, but a number from the existing list followed by *bis*, *ter*, *quar*, etc. In this dissertation, artefacts attributed by Beazley are cited according to the painter, the vase number, and the references in whichever of these works the vase is listed. The vase number will be listed after the page number of the earliest of these works in which it appears. For example the Princeton Painter's neck-amphorae, London B212, would be referenced as "ABV 297.1, *Para 129*, Add^2 78" indicating that the vessel is first listed on page 297 in ABV and is number one in his master list, and that the vase is cited in *Para* on page 129 and in Add^2 on page 78. On the other hand, the neck amphora Leningrad 162, which does NOT appear in ABV is referenced as "*Para* 130.1*bis*, Add^2 78". Note that the number on the master list (1 *bis*) is after the page number in *Para*.

In addition to Beazley's lists there are some other specialised lists of mythological iconography. These are referenced only if the respective myth is discussed for some reason. Three are cited in this dissertation. First, Brommer [1973]'s list of Greek heroic iconography, *Vasenlisten zur Griechischen Heldensage* (*Vasenlisten*)is referenced by mythical figure, page number and vase number. Second, the *Lexicon Iconographicum Mythologiae Classicae*[18] (*LIMC*) is referenced by volume number, mythical figure and vase number. Finally, von Bothmer [1957]'s *Amazons in Greek Art* (*Amazons*) is referenced by page and vase number.

Finally, a very important resource for photographs and descriptions of vases is the *Corpus Vasorum Antiquorum* (*CVA*) which is an occasional pub-

[15]Beazley [1968].
[16]Beazley [1971].
[17]Carpenter [1989].
[18]Ackerman and Gisler [1981-].

lication conducted by individual museums under the auspices of the Union Academique Internationale. The aim of of *CVA*, which has been an ongoing concern since 1919, is to make images of all Greek vases available to scholars around the world. *CVA* fascicles are released sporadically when a museum makes their collection available in this way. *CVA* photographs are referenced in this dissertation by the country, the volume number, the museum and the fascicle number, followed by the plate number and the page number.

1.2 Aims and Scope

The aim of the study is to demonstrate the principle that pattern recognition may be used as a tool to aid in the stylistic attribution of Attic-vase paintings according to the categories established by Sir John Beazley. The dissertation both presents a set of methods to aid in the classification of vase-paintings according to style, and demonstrates them on the *oeuvre* of the Princeton Painter, an Attic black-figure painter active in the third quarter of the sixth century BCE. The topic is potentially very broad and has been delimited in order that it may be treated in sufficient depth. In addition, and more informally to avoid any misconceptions about the aims and goals of this project, I will first explain what the project does NOT attempt to achieve.

First, the aim of the study is considerably more modest than the implementation of a digital connoisseur to replace the art historian at the attribution of art. Such a task is monumental and this study is just a small part of a nascent discipline. Traditionally, attribution in art history has typically been based on a combination of the careful study of a few salient features (such as Morelli's *grundformen*) combined with the intuition of the art historian. Recently, however, numerous scientific methods have been proposed in the attribution of art (discussed in more detail in 1.4.2), none of which have claimed to be "silver bullet" automatic attribution techniques. Rather they are presented as techniques that may be used by the art historian as an aid in the attribution. This dissertation is in keeping with this spirit by provided techniques that are not meant to be used in isolation. However, particular care has been taken to illustrate the efficacy of each technique and the methods by which these techniques may be combined and used in concert with the intuition of an art historian are suggested in 6.2.2.4.

Secondly, the present study does not claim to be a comprehensive study of the *oeuvre* of the Princeton Painter. While the vases of the Princeton painter are discussed, such discussion aims to introduce the reader to the painter serving as a model for the techniques presented, and to explicitly state the categories by which the computer-aided classifications will be based. This necessarily means that the very interesting issue of the Princeton Painter's relationship with other painters in the Princeton Group has not been dealt with in any detail. However, a brief discussion of the Princeton Painter's style, his stylistic relationship with contemporary painters and a tentative chronology of his works is provided in the Appendix.

Finally, the purpose of the dissertation is neither to prove or justify the practice of connoisseurship by means of statistics nor to prove with any rigour that Beazley was correct in his attributions. Instead, a fundamental assumption of this study is that connoisseurship is a valuable tool in the archaeologist's kit, and that a computer aided method to aid in this process will be of use to the discipline. Thus, I will assume for the purposes of this dissertation that Beazley's classifications are correct and aim to show how, on the basis of this assumption, computers may be used to simplify the way in which such attributions are made. Whether or not this assumption is correct is irrelevant to this outcome of the dissertation: the intention is simply to make the first step towards the automation of a process that is already used extensively by archaeologists studying Attic black-figure.

Nevertheless, a defence of Beazley and connoisseurship is supplied in the next section (1.3.2), and furthermore, the efficacy of the proposed techniques is demonstrated on modern Japanese prints (4.2.5) where the attributions of the objects are not in doubt and in chapter 5 it is shown that vases may be attributed by potting shape, even in cases where it is clear that potting shape was not a criterion used by Beazley.[19] It must be stressed however, that these are not claims to a scientific proof of Beazley's method, but may be viewed as supporting evidence. It is the belief of the author that scientific proof or disproof of Beazley's method, or of some of his attributions, may be possible, but not at this stage. The first step should be, instead, to provide some more formal method to conduct the attributions before these formal methods may be placed under close scientific scrutiny. Again, it is not the aim of this study to provide all of these formal methods, but by example, to illustrate that they may be achieved.

1.3 Motivation

1.3.1 Rationale

This study is very narrow and may appear to be disjoint from current trends in archaeology and, as such, requires special motivation to explain it's rele-

[19]In the first instance, Beazley did not use shape much. Most of his attributions were conducted on the basis of photographs.

vant to the central questions of the discipline. The classification of artefacts is one of the core methods of archaeology. The practice, introduced in the mid 18th century, had established itself as a primary facet of archaeological investigation by the mid 20th century. New archaeological methodologies that gained popularity in the latter half of the 20th centuries stress the importance of social, political and anthropological theories in the interpretation of artefacts, but classification and typology are still central to the field today. The process of classification is laborious and relies on time-consuming descriptions. Attribution may be seen as one of the methods of description and classification, albeit one that usually takes place outside of the excavation site and sometimes only just before the ceramics go into the market. Because it is so specialised, attribution is often not undertaken by excavators (and in fact, many of the pieces in the market were not excavated, but looted).

Stylistic classification allows for finer dating than other methods since, if a painter's style is well-understood to the point that his works may be placed in a chronological sequence, it is possible to make a good conjecture as to where any new pieces fit into that sequence. Furthermore, style based classification is very important for the reconstruction of vases from the mass of otherwise undifferentiated sherds. As it is extremely unlikely that any two sherds painted by different hands should come from the same pot, classification according to style places very tight constraints on which sherds should be considered for any single reconstruction. Beazley on occasion, mainly on stylistic considerations, correctly matched sherds in disparate collections [Boardman, 2006, p.134]. Reconstruction of whole vases from sherds is useful for Classical scholars because, as has been pointed out, the scenes offer a valuable supplement to the literary record.

Unfortunately, Beazley's typology is based largely on a methodology that has not been explicitly published (it is usually transmitted from teacher to pupil),[20] is expensive and arcane, and is specialised to the point that sherds often are attributed by specialists. This means they often have to be examined by an expert before they can be properly dated, used in vase reconstruction, or used by Classical scholars as visual evidence. The rate at which new connoisseurs of Attic pottery enter the field is slowing down rapidly, as other avenues of research are, quite rightly, gaining in popularity.[21] A semi-

[20]This specifically applies to Beazley's method. Other connoisseurs like Morelli [1890], Berenson [1954] and Friedländer [1969] have published their methodologies. Boardman [2006, p.131] argues that Beazley's earliest articles do explain his technique (at least implicitly).

[21]A statement like this is difficult to prove. However a powerful piece of anecdotal evidence is that at Oxford, Beazley's department and the largest department of Classical

automated process of attribution would allow scholars with only a limited understanding of this arcane subject to engage in the practice. In particular, this would mean that archaeologists in the field will be able to attribute sherds as they are uncovered, facilitating both the process of reconstructing whole artefacts, and also the process of dating them. Yet, while many computer-aided techniques for automated or semi-automated artefact classification have been proposed (1.4.2), only two have dealt with the issue of style.

Finally, this study lays the foundation for further reseach into scientific attribution which, if it is achievable, might put an end to the practice of art dealers spuriously attributing works so as to maximise the price of the artefacts. Furthermore, when attribution is no longer seen as the province of learned academics and connoisseurs, but rather the province of the scientist and the computer, perhaps some of the prestige of painted Greek vases will be diminished together with the consequent illicit practices of the art market that is driven by demand for these artefacts.

1.3.2 A defence of Connoisseurship

Narrow stylistic studies of individual painters are more and more being considered inappropriate areas for research, not only in the study of Renaissance and later art, but increasingly in archaeological studies as well. The reason is that connoisseurship has come under severe attack from a number of quarters and for several reasons, some of which will be discussed. However, the tasks of attribution and authentication are important in art historical studies for a variety of reasons. In particular, as has already been explained above, it provides valuable information for archaeological studies of these artefacts.

While this dissertation is not attempting to prove the authenticity of Beazley's method of attribution, it is predicated on it. Therefore it is necessary to provide some defence of connoisseurship. This defence takes the form both of a rebuttal of the unwarranted attacks and an attempt to explain how those that have any merit are either not applicable to the study of vase painting, or are addressed in some way by the present study. The defence is broken into two parts. The first deals with connoisseurship in general (i.e. not limited to archaeology) and with criticisms emanating from within the art historical community. The second deals specifically with criticisms of connoisseurship applied to vase painting and with criticisms emanating from within the ar-

art history in the English speaking, there are no dissertations in the pipeline that deal with attribution and I have been informed that such a proposal is unlikely to be accepted today.

chaeological community - specifically from Michael Vickers and David Gill whose hypothesis is treated in considerable detail.

1.3.2.1 Connoisseurship in General

Some view the discipline of art historical studies as comprising two broad approaches: the new and old art histories. The latter is concerned with interpreting art within its sociohistorical context and the reciprocal interpretation of a society through its art. The old art history, on the other hand, is concerned with the analysis of the objects themselves. This usually takes the form of both connoisseurship one the one hand, and iconography and iconology on the other.[22] The main claim against old art history is that it is outdated and unnecessary. In particular, it is often argued that such narrow studies are ends in themselves, being of little service to other areas of study. This is not an entirely fair criticism, however, as narrow studies are very useful as reference works. For example, anyone using a painting as evidence needs to be able to interpret it and may often require the help of a narrow monograph on the painter or on the iconography. More importantly, this criticism does not hold for the study of Attic vase-painting, since the study of vase-painters has a definite benefit for archaeologists - its usefulness for dating and sherd reconstruction. This is all the more true given that, unlike European art of the Renaissance and later, vase paintings are seldom signed.[23]

A second form of criticism commonly used against connoisseurship is to illustrate the failings of some connoisseurs and cite this as evidence that the techniques are flawed. For example, Bernard Berenson, the flamboyant critic and disciple of Giovanni Morelli was often criticised for his use of purely subjective criteria for evaluating objects and for his mistaken attributions. Examples of Berenson's failures are sometimes used as evidence that Morelli's technique, which supposedly underlies Berenson's own, is fundamentally flawed. Of course, these criticisms cannot be taken as evidence against the whole notion of style - based attribution, but are criticisms of individuals who either misapply the method, or whose own methods are flawed. In particular, Berenson did not simply use Morelli's "scientific" method, but a large part of his work was subjective. Therefore an objective method of attribution should not be subject to criticisms derived from the failings of Berenson.

A third criticism is that Morelli's technique soon became a victim of its

[22]Iconology and iconography are the study of interpreting works of art in isolation.

[23]Boardman [2006, p.128] claims that 40 of 900 artists have extant signatures and estimates that less than 1% of the surviving pots have signatures.

own success. As many connoisseurs used Morelli's method to detect forgeries, it became easy for forgers simply to concentrate on those features that Morelli considered important. Thus, because the forgers and the detectives were working from the same manual, so to speak, it was easier for forgers to remain ahead of the game, by pre-empting the connoisseurs. This criticism is only valid when Morelli's technique is used to determine the authenticity of a signed work of art rather than to determine the authorship of an unsigned one. In the case of vase-painting the method is not used for authentication at all. Instead the techniques of attribution are used to determine authorship of unsigned vases. Modern forgeries on the other hand are not only easy to spot, but can be detected by other methods, such as thermoluminescence.

A final, and a very important criticism, is that connoisseurship has been irrevocably tainted with the sordid affairs of the art market. Morelli was hated by those who felt that connoisseurs were really the puppets of the art dealers and collectors. To some extent, the behaviour of connoisseurs like Berenson has done little to dispel this notion - Berenson himself was quick to sell his services to art dealers, and it is very difficult to believe that his pronouncements were entirely unbiased. Worse still, his opinions were perceived to carry the weight of science when, as Berenson [1954] himself admits, a considerable part of his attribution was based on subjective criteria.

In the case of Attic black figure, it is clear that there is a considerable illicit trade in artefacts that is being fuelled by unscrupulous art dealers who do not question the provenience of the wares that they sell to collectors whose tastes are shaped by the connoisseurs who treat the artists who created these vessels in the same way they would masters of the Renaissance. The notion is false, vases were largely utility vessels created by humble craftsmen, some of whom were able to become quite wealthy, and some of whose vases may well have been valued quite highly. One consequence of this high esteem of these vessels is that the prices are now so high that small museums cannot afford them and many of them land in the hands of wealthy collectors and are effectively removed from the archaeological record. Secondly, because of the illicit trade in antiquities, the artefacts are orphanned from their archaeological contexts,[24] severely limiting our ability to interpret them. However, this is not a convincing reason to abandon connoisseurship, but rather to put in place measures to ameliorate the situation.

[24] In fact, this in itself has been used as a critique of connoisseurship of Attic black figure - that the context of so many of the artefacts is lost, our baseline data about the objects is very scant.

1.3.2.2 The Vickers-Gill Hypothesis

More specifically, in the field of Attic vase painting, there has been a concerted attack on the orthodoxy of connoisseurship by Michael Vickers and David Gill (hereon V&G). V&G make a number of valuable observations about the state of the discipline and make some strong claims about the validity of connoisseurship in Attic black-figure. I present a brief summary of their argument before visiting the details. V&G argue that Greek vase painting is overvalued today and that this overvaluation is due to a misconception about the value of the works in antiquity. As a consequence of overvaluing such pottery today, we overstate the value placed on the originality of these works in antiquity, and consequently ignore the evidence that ceramic vessels were direct copies of metal vessels. V&G further suggest that the artistic personalities identified by Beazley and others are not the painters of the vases, but the people who made the metal vessels from which the vases were copied. Finally, V&G illustrate the damage that connoisseurship has done to the study of vase paintings. What follows in this sub-chapter is a detailed critique of the major points raised by Vickers and Gill. It is the opinion of the author that, while many of their hypotheses either fail or have no little evidence, there are some points that are of considerable importance. The rebuttals presented here are not original, since many commentators have responded to Vickers and Gill's attacks.

A point repeatedly stressed by Vickers and Gill [Vickers, 1983, 1985, 1990, Vickers and Gill, 1994], is that Greek ceramic art is overvalued today, and that it was practically worthless in antiquity. There are a number of consequences of this overvaluation, the first of which is that it drives the demand for these artefacts and this in turn drives the illegal art trade and worse still, the purchase of artefacts for private collections in which they become lost to scholarship. Vickers cites evidence that the better vases sold for less than 3 *drachma* while precious metal wares would sell for several hundred times this value. By comparing the relative value of gold to silver in the modern world, and comparing the relative value of ceramics to silver in the ancient, they arrive at a value of £1-80 to the *drachma*. Obviously, ancient ceramics were not expensive. Nevertheless others have argued that Vickers and Gill have severely exaggerated this comparison. First, Boardman [1996] points out that such direct comparisons do not reflect the structural differences between ancient and modern economies; the average daily wage in Ancient Athens was 1 *drachma*, hardly the equivalent of £1-80. In any event, modern tastes are not dictated by the value of objects in the past. A trivial example may illustrate this: Beatles records would have sold in the 1960s for exactly the same price as other records, but their value today is much higher than that for other less collectable artists. While V&G have overstated their case, they have rightly brought to the attention of Classicists that these wares were not valuable in antiquity.

Perhaps the reason that V&G exaggerate the worthlessness of vases is because they argue that we should consequently not imbue in these works any artistic originality. The argument runs as follows [Vickers and Gill, 1994, ch4]. The notion that originality could imbue humble crafts with the dignity of art is only as old as the age of Enlightenment (roughly the Eighteenth century), and the concept would have been totally foreign to the Greek artists. For these humble craftsmen only the ornamental value of the decoration was of significance, not the original compositions. In other words, the value of the vases, in antiquity, lay in the beauty of their decoration, rather than in their artistic innovation. This lays the foundation for their primary attack on the methodology of connoisseurship: their hypothesis that Attic vases were copies or skeuomorphs,[25] of vessels made in precious metals, and at best attribution of a ceramic vessel tells us about the designer of the metalware from which it was copied, and not about the maker of the ceramic vessel itself. There are many threads to this argument and treating them all in detail is redundant, since it has already been done by others. Instead what follows is a summary and a critique of their most important points.

Having established that vases were worth considerably less in the ancient world than today, and having questioned the notion that originality was considered an artistic virtue by the craftsmen, V&G proceed to point out some unsolved problems in attribution studies and attempt to show that their thesis solves all of these. Problem one is that one of the conventions by which artists signing their vases does not appear to be systematic. The phrase "*X egraphsen*" literarally means "X painted/drew (me)", and "*Y epoiesen*" means "Y made (me)". The traditional view is that latter is considered to refer either to the owner of the workshop,[26] the potter or to both (the potter could also be the owner of the workshop). V&G point out that there are problems with this interpretation and instead suggest that this dual signature reflected a practice that they believe was common amongst metalsmiths. Their first evidence is that there exist vases by obviously different hands that have the same signature. The most famous is a cup attributed to the Triptolemus painter that is inscribed "*Douris egraphsen*". This has traditionally been interpreted either as a boast by the Triptolemus Painter that he is as good as Douris, an attempt by the former to forge a Douris, or perhaps even

[25] The conventional archaeological term for objects that contain elements of design more appropriate to another class of objects of which they are imitations.

[26] A position advocated by Cook [1971].

an homage. Secondly, they point out that the term *kerameusen* is a more natural one to describe a potter, yet it is used on only one extant vase which V&G claim shows no resemblance to metalwork. Third, they cite Atheneus who claims to have seen a silver cup bearing the inscription "Gramma Parrasioi Techna Muos" (drawing by Parrhasius, art by Mys)[27] and Pliny (*Natural History* Book 35 line 68)[28] who claims that Parrhasius' sketches on parchment were subsequently used by many painters. This they believe is evidence that (a) parchments were used to draw preliminary sketches, (b) the parchment was valuable enough to be kept and copied, and (c) the practice on metalwork was to incise signatures indicating a division of labour between draftsman and craftsman. On this basis they conclude that the "*epoiesen/egraphsen*" signatures on pottery should not be seen as indicating a division of labour between the potter and painter, but are faithful copies of the inscriptions on the parchments that they copied.

Cook [1987b] published a long critique to Vickers [1983] and Vickers [1985] but the issues he raised were not addressed in Vickers and Gill [1994]. Briefly, Cook notes that if parchments were so rare and expensive, and the copying from them so faithful, then one would expect the extant pots to reflect less variety in the scenes than they do. Furthermore, the distribution of *egraphsen/poise* among the potters suggests that stylistically most vase painters use exclusively the designs of one silversmith, which would be surprising given the worth of these parchments. Finally, Cook points out that the *epoiese/n* signature is found on undecorated ceramic pots, indicating that the craftsman was not working to a design, and that "*Echsekias egraphse kapoiese me*" (Exekias painted and made me) appears on two elaborate amphorae[29] where clearly no division of labour is needed. These points do much to undermine V&G's claims.

A second problem cited and addressed by V&G is that in the middle of the last quarter of the sixth century BCE, the dominance of the black-figure style is challenged and eventually brought to an end by a new technique: red-figure (1.1.1), which remains the dominant technique until 320BCE. V&G argue that interests in artistic problems could not have prompted the shift from black- to red-figure because red-figure is so much harder to render, and they provide an explanation which is consistant with their theory that vases are direct and faithful copies of metalworks. First, they claim that the colours on vases reflect the colours on metalworks. For V&G black represents tarnished

[27] Parrhasius is perhaps the most celebrated artist from the Greek world (Pliny *Natural History* Book 35 line 68), and Mys one of the most celebrated silversmiths (Pliny *Natural History* Book 33 line 155).

[28] [Jones, 1951]

[29] Berlin 1720 *ABV* 143.1, *Para* 59, and Vatican 344 *ABV* 145.13 & 686, *Para* 60.

silver (i.e. silver sulphate), white is ivory, added red (which V&G call purple) is copper and reserved is gold). The suggestion is then made that the reason for the change from black- to red-figure is that workers in silver who had since used gold overlays with the figures cut out, started instead to use gold figures as overlays because this would require considerably less gold. The potters, who faithfully copied from metalwork, simply mimicked this trend. Thus, red figure, in which the background is black and the figures are reserved would mimic silver pots with gold leaf for the figures whereas black-figure in which the background is reserved, mimics gold leaf background with silver male figures and ivory females. V&G also claim that the disappearance of red-figure in the late fourth century BCE was due to the Greek elite developing a taste for near-Eastern style engraving in metalware after the Macedonian conquests, and consequently the potters duly started to copy this medium and the result was Megarian ware, an example of which is shown in figure 1.6.[30]

This line of reasoning has been thoroughly addressed by Cook [1987b] who points out that after Homer, terms with the root "*argur-*" (silver-) are most often used to describe light coloured objects, such as the moon in a fragment of Sappho [Cox, 1925, 1.4] which brings into doubt V&G's argument that silver was meant to be black. In addition, Cook points out that Pliny, in his long exposition on chasing silverware[31] does not mention any change from tarnished to polished silver during the Hellenistic period. Boardman [1996] has pointed out other problems. For example, V&G's theory also has difficulty explaining why there is considerable continuity in ceramic shapes that long predates the mass production of metal vessels, and Simon [1996] points out, in particular, that kantharoi have an extremely long history that goes back to Troy. V&G's explanation for the change from black- to red-figure and the disappearance of red-figure is also problematic. They do not explain, first of all, why it should take so long (since the 7th century) for metalworkers to realise that gold backgrounds are more expensive than gold figures, and in particular why this economic motive should express itself at a time of relative abundance of wealth in Attica. Furthermore, V&G's explanation for the disappearance of red-figure is slightly anachronistic since red-figure's immediate successor is not Magarian ware[32] which, from figure 1.6 clearly does mimic metalware. Instead, it is red-slip ware which immediately replaces red-figure and red-slip ware does not obviously imitate metalware [Cook,

[30]Edgar Lowen catalogue item 5297; image from Edgar Lowen Antiques: http://www.edgarlowen.com.

[31]*Natural History* Book 33: [Jones, 1951]

[32]Rotroff [1978] finds evidence that Megarian ware may be an innovation from as late as 225 BCE, although conventionally this is thought to be around 275-250 BCE.

Figure 1.6: An example of Megarian pottery. It is clear from the relief on the side that this pot is a skeuomorph of engraved metalwork.

1987b]. Thus V&G's argument that red-figure comes to an end because of a change in the style of metalwork is untenable.

A final thread of V&G's argument is that there is a great deal of illiteracy evident in Greek vase-painting with frequent nonsense inscriptions and a lack of orthography in the spelling of names. This they cite as evidence that the works were copied, the idea being that the painters being lower-class, were more likely to be illiterate and thus mis-copied the inscriptions from the metalworks they were imitating. To some degree this undermines their argument that the *epoiesen/egraphsen* inscriptions represent a faithful transcription of the signatures on metalworks. More problematic, however, is that it undermines their claim that Attic vessels were copied from metal works with such great fidelity that the subtleties of the style of the metalsmith was evident to such a degree in the ceramic "copies" that attribution was impossible. If this had been the case, then surely one would have expected the copying of gross features like letters to be equally faithfully copied.

I should like to stress at this point what is required for V&G's argument to undermine the practice of connoisseurship. They require not only that Attic pottery are skeuomorphs of metalworks, but that the copying is so faithful that we cannot infer anything about the style of the painter who decorated the pots, only the style of the metalworker from whom he copied. On this point, there is considerable evidence that vase-painters have copied from other media, but it is unusual for this copying to be slavish. This may be illustrated with an example of a motif from another medium that was very influential in Attic black figure but which allows for a great deal of variety in expression. This is the east pediment of the Siphnian treasury at Delphi which depicts Herakles' and Apollo struggling for the tripod (figure 1.7).[33] The treasury was erected in 525 BCE, and immediately afterward, the popularity of scenes of Herakles and the tripod not only increased, but the composition of the scene changed to match that of the treasury [von Bothmer, 1977]. However, even within this there is very considerable variation, not only in style, but also in composition, as is evident from figures 1.8(a)[34] and 1.8(b).[35] Even in cases where the compositional template on two Attic vases are exactly the same, there is considerable stylistic variation. A further example of a scene that has many examples in black-figure but which are almost exactly alike in composition is that of the chariot wheeling around.

[33]Photograph by Ellen Brundige 2005 (c) Wikipedia Commons.

[34]Madrid Inv.10913 by the Madrid Painter *ABV* 329.9 *Add*^{2} 89. Photograph by Marie-Lan Nguyen (c) Wikipedia Commons.

[35]London G180 by the Siren Painter *ARV*^{2} 289.2, 1642 *Add*^{2} 210. Photograph (c) Wikipedia Commons.

Figures 1.9(a)[36] and 1.9(b)[37] illustrate two different examples of the same scene, attributed respectively to the manner of the Princeton Painter and Group E by Beazley. The two images not only show some variation, but the stylistic features are quite different. Notice, for example, the incisions used to indicate the horse anatomy, such as the hook at the end of the eye of the first horse from the left on 1.9(a) which is absent on 1.9(b), the inverted Y shape reign on the horses with frontal faces on 1.9(a) compared with the single horizontal line around the muzzle in 1.9(b), the three arc segments on the thigh of the horses in 1.9(a) versus four in 1.9(b). These stylistic elements illustrate that even the most faithful reproductions in black figure do not preserve these stylistic features of the original.

To sum up this complicated debate, V&G maintain that Athenian ceramics were very cheap and that the potters and painters who fashioned them were not considered to be artists by the ancients, that the vessels they produced were not innovative and revealed no clues about the identity of the makers, since they were skeuomorphs of metalworks which they had copied so faithfully that the hand of the ceramicist was no longer discernable in the final product. However, V&G's arguments are not widely accepted amongst Classical art historians. It is generally accepted that, while there may be some exaggeration, V&G are correct that ceramics were very cheap and abundant. Few scholars, even those sympathetic to their cause (for example Elia [1996] and McClellan [1996]), however, are willing to buy V&G wholesale. In particular, even if there is considerable copying from metalwork (and there is certainly evidence that wares in different media did influence each other) there is little chance that it was so faithful that the hand of the ceramicist is not evident in his work. Arguments about the value of pottery in antiquity and the status of these works as art are of little consequence to the attribution of the works, particularly in the case of computer based attribution where aesthetic concerns are of no importance.

On the other hand, it is important to read the implicit (and not particularly subtle) subtext of V&G's attack on the establishment: that connoisseurship is doing a great deal of damage to the study of Greek archaeology. David Gill has pointed out, regarding Cycladic figurines:

> "Archaeological sites around the Mediterranean and elsewhere are suffering major damage due to systematic and illicit excavations in order to supply the needs of the antiquities market. This activity in turn feeds the appetites of the museums and the private collectors who

[36] Munich 1376, Manner of the Princeton Painter, *ABV* 300.12 *Para* 130, *CVA* Germany 105, Munich I 12-13, pl. 12.3.

[37] Louvre CP10659 by Group E *ABV* 138.69 *Add*^{2} 37 *CVA* France 11 Paris 3 pl 141.1-4.

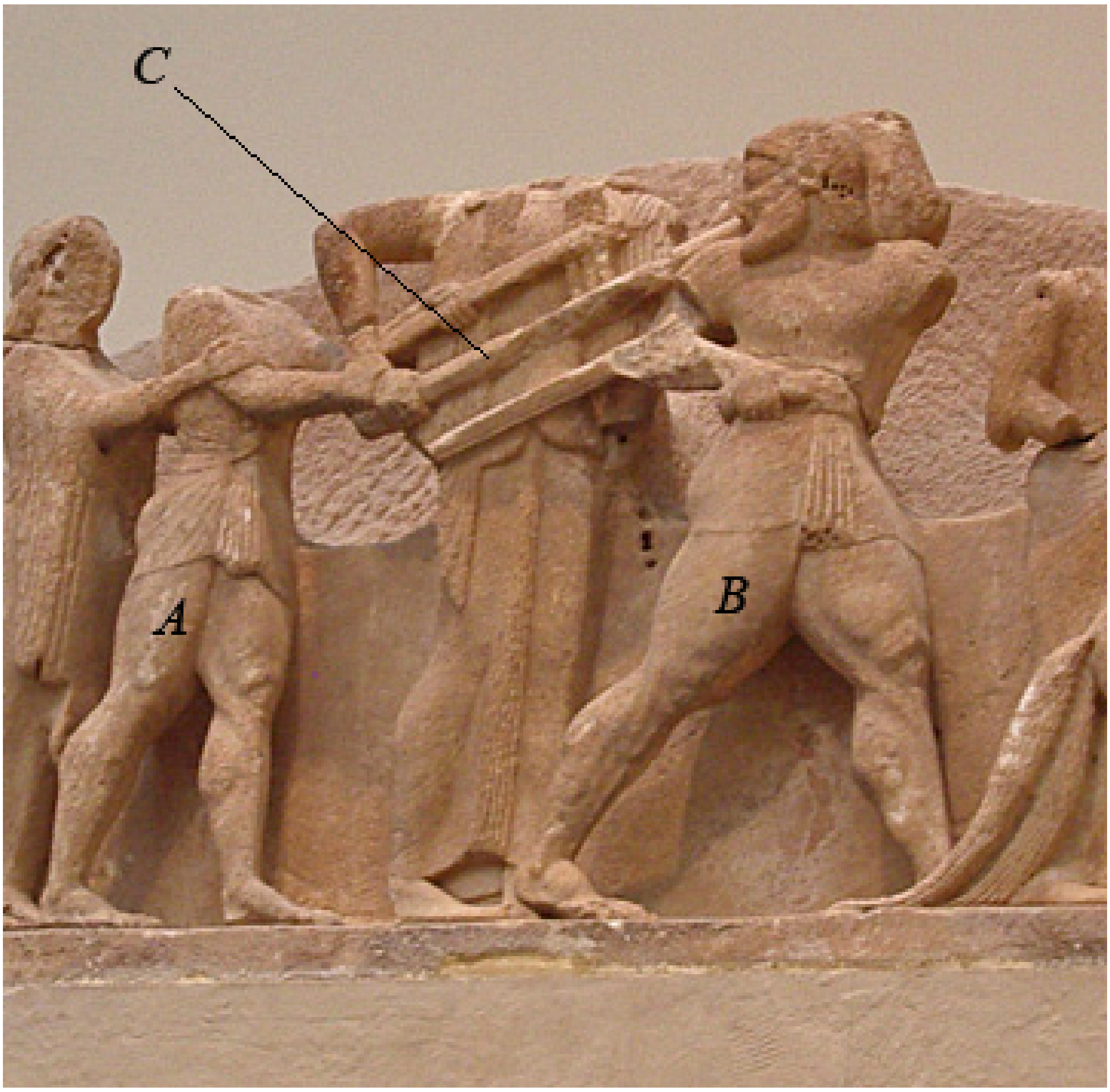
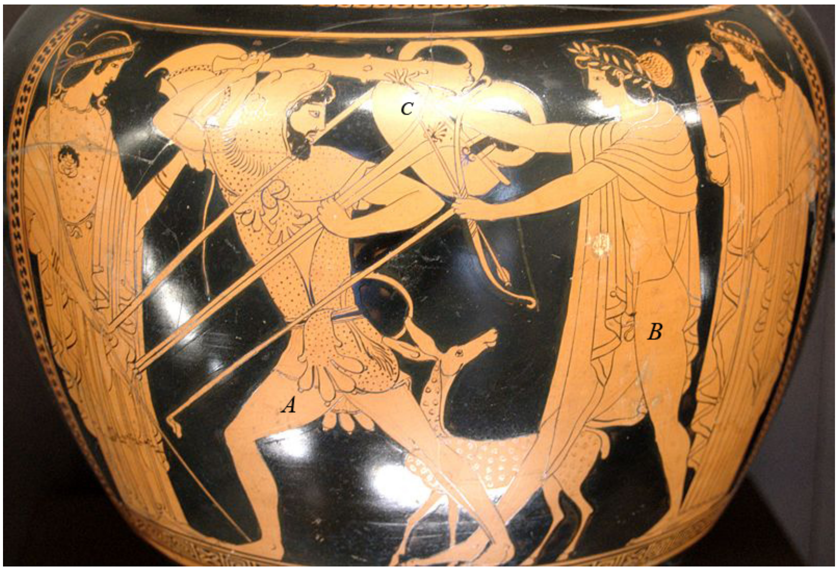


Figure 1.7: The east pediment of the Siphnian treasury at Delphi depicting Apollo and Herakles fighting over the Delphic tripod. A) Apollo, B) Herakles and C) the tripod.

(a)



(b)

Figure 1.8: Two vases depicting Herakles and Apollo fighting over the tripod. Both are thought to be influenced by the Siphnian treasury. (a): A) Apollo B) Herakles C) the tripod; (b) A)Herakles B) Apollo C) the tripod.

(a)

(b)

Figure 1.9: Two vases depicting a chariot wheeling around, (a) in the Manner of the Princeton Painter and (b) by Group E. The template of both scenes is so similar as to suggest that one is a copy of the other or, more likely, that they are copied from the same source. However, the stylistic features are very different.

> are willing to buy. A case study on marble Cycladic figures has suggested that some 85% of the funerary record of the Early Bronze Age Cyclades may have been lost through this unscientific search for figurines"[Gill and Chippendale, 1993, p.625]

While the situtation in the study of Attic vase painting is not as serious, the same problems are apparent. First, the high prices paid for these artefacts leads not only to looting, but to a very rushed process of transferral from the site of excavation to the marketplace, often by illicit means. This means that the archaeological context is, for the most part lost, and with it a great deal of very valuable scientific information. In addition, the practice of attributing vases according to artists, as though they were Renaissance masterpieces, serves implicitly to justify their demand from private collectors. This demand pushes up the prices for the artefacts which in turn moves them out of the reach of many museums and into the collections of individuals where they are lost to modern scholarship. Clearly, the situtation is intolerable and much needs to be done to rescue the discipline from these practices. But, given the lack of provenience and context on so many of these vessels, stylistic analysis should be seen as a valuable tool, given the insight it can provide regarding not only the dating of the vessels but for interpretation and reconstruction of sherds. Dispensing with a tool as valuable as attribution in order to save an archaeological record that is already in tatters will not ameliorate the problem. Instead, we propose that the following steps would be more helpful. (1) Countries should close loopholes in their laws that allow artefacts without provenience to be sold to private collectors. (2) There should be stringent conditions governing the sale of cultural artefacts including (a) that these artefacts should be available to interested scholars, (b) they be sold only on condition that they have already been properly documented in site reports and (c) that the objects should be photographed and captured using the techniques discussed in 6.3.2 and that these images be made available to the academic community through repositories like the Beazley archive.

1.4 Literature Review

The present study does not sit perfectly within the purview of any single discipline and a review of cognate literature is a multi faceted task drawing, as it does, upon similar works from a number of different disciplines. In particular, the study borrows both from connoisseurship and from statistical pattern recognition. Three particular areas of study within these disciplines are both of particular relevance to this thesis and have a substantial literature. These are the three headings under which research cognate to this study will be surveyed: the study of the Princeton Painter and related artists, computer-aided analysis of artistic style, and machine learning analysis of Attic ceramics. To the last category belong two dissertations concerned with the analysis of style in Attic black-figure vase painting, and these two have been dealt with in more detail in 1.4.3.

1.4.1 The Princeton Painter and similar artists

The first publication recognising the Princeton Painter as a distinct artistic personality was Beazley's *BSA* paper on Attic Black-figure [Beazley, 1932, pp.17-18] in which he identified for the first time a number of previously unknown black-figure artists. Under the Princeton Painter he listed only two vases.[38] He would add another 23 of his own attributions in *ABV* and a further 4 in *Para*. In *ABV* Beazley also included 2 attributions by other scholars, one by Smith[39] and one by von Bothmer, who had included the vase in his PhD dissertation on Amazons.[40] There have also been a number of attributions to the Princeton Painter by people other than Beazley. Elke Böhr [1982], in her monograph on the Swing Painter made the tentative attribution of Stuttgart 65/1 to the Princeton Painter, as well a number of vases to his manner. von Bothmer attributed a neck amphora in Geneva [Chamay and von Bothmer, 1987] and another amphora in New York, a panathanaic from the Norbert Schimmel collection, later endowed to the museum in 1989 [Chamay and von Bothmer, 1987, p.64]. The new New York attributions were published together with a new attribution[41] by Mary B Moore in a recent edition of the *Metropolitan Museum Journal* [Moore, 2007]. In the same article, Moore also confirms attributions to this painter by Heide Momsen[42] and Elke Böhr[43] Moore [1975] had also published some fragments of a column krater in Samothrace to the Princeton Painter.[44] This attribution is of some interest because it bears the mark of the potter: *Werxekleides epoiesen*. The name itself has received some scholarly attention because of the unusual lettering [Boegehold, 1983] [Boardman, 1983].

Although there are a number of attributions to the Princeton Painter's hand, he has not been the subject of much scholarly interest. Only three pub-

[38]Boulogne 4 **R4** and Princeton 169 **R7**.

[39]New York 56.171.9 **M10** attributed by Smith [1945, p.463].

[40]Cincinnati 1884.213, **M13** *Amazons* 38.1.

[41]New York 1911.11.2 [Moore, 2007, pp.21-28].

[42]Bochum S 1205(type B amphora).

[43]Munich 1385 (type B amphora) and an amphora once in Summa Galleries, Beverly Hills (panathanaic).

[44]The attribution had been suggested by Ellen Davis (according to Chamay and von Bothmer [1987, p.62]), but Moore provided justification for this attribution.

lications concern the Princeton Painter's works as their primary subject matter. The first of these is von Bothmer's publication of a pseudo-Panathanaic amphora in New York in which he tries to interpret the unusual iconography of the vase von Bothmer [1954].[45] The second article is a substantial discussion of a neck-amphora in Geneva [Chamay and von Bothmer, 1987]. The article is in two parts. The first, by Jaques Chamay, discusses the myth depicted on the obverse - The rape of Kassandra. The second, by von Bothmer, is a discussion of the Princeton Painter's favourite themes and his subsidiary decorations, although it contains only a limited discussion of his style. In it, von Bothmer states some of the difficulties of establishing a chronology of the Princeton Painter's work and studiously avoids this task. Finally, Mary B Moore discusses 5 vases by the Princeton painter in the Metropolitan Museum of Art [Moore, 2007]. In the article Moore examines the respective vases in detail, including a brief look at the ornament, the style and the provenience. In addition, Moore confirms Böhr's 5 attributions. However, as von Bothmer, Moore does not treat the difficult issue of the Princeton Painter's chronology.

For the purposes of this dissertation, it is important to define who's attributions are to be accepted, as these are the training samples for the learning algorithms developed in this thesis. In all cases, including those of the artists used for comparison[46] we will only accept vases attributed by Beazley or attributions accepted by Beazley and included in his lists. This is not meant as a rejection of the attributions of Moore, Böhr and others, but rather that since the ostensible aim of this study is to prove that a classifier can be trained to classify in Beazley's manner, those works that do not officially carry his stamp of approval cannot be included.

1.4.2 Computer Aided Style-Analysis

In the early seventies, George Stiny and James Gips proposed a system of specification by which designs are described by a formal language called a

[45] The obverse depicts a scene reminiscent of the Birth of Athena in which Zeus appears seated surrounded by gods with a fully armed Athena bursting from his head. This scene is unusual in that Athena does not appear, but the template is exactly the same and difficult to interpret. The reverse shows Athena as is customary on Panathanaic amphorae. However, on the New York Amphora, Athena does not appear between the customary two columns surmounted by cocks, but is framed on the left by an *auletos*, facing right, and a flaming alter, and on the right by a maiden, facing left, who carries a fillet in her hand and on her head carries the folded *peplos*, which served as a ritual gift to Athena presented during the Panathanaic festival.

[46] Exekias and Group E in Chapter 3, Exekias, Group E, the Painter of Berlin 1686 and the Swing Painter in Chapter 4, and Exekias, Amasis and Andokides in Chapter 5.

shape grammar.[47] Shape grammars contain production rules and a set of primitives, which are generally shapes or classes of shapes. These may be used to generate more paintings or designs. Style can be specified by both the primitives (what sorts of shapes are likely to be found in a given style) and by the production rules, which constrain the ways in which these primitives may be combined according to the rules of that particular style. Using the system, Stiny formally described the style of Chinese lattice designs [Stiny, 1977] and Stiny, Gips and Mitchell described Palladio's villas[Stiny and Mitchell, 1978, Gips and Stiny, 1978]. The first attempt at computer-aided analysis of the style of individual painters was by Russel Kirsch, a pioneer of digital image processing.[48] Further to the purely mathematical formalisms of Gips and Stiny, Kirsch formulated a set of algorithms by which a computer could replicate the styles of the abstract artists Richard Diebenkorn [Kirsch and Kirsch, 1986]and Joan Miro [Kirsch and Kirsch, 1988, pp.441-444], and also North American Indian petroglyphs [Kirsch, 1997, 1998]. Areas in which shape-grammars have been used formally to describe non-architectural art include, among many others, Knight's analysis of meander patterns on Greek geometric pottery [Knight, 1986] and of de Stijl paintings [Knight, 1989].

Despite its use on a wide range of styles, there are a number of reasons why shape grammars are not appropriate as the basis for an attribution study of artistic style. First, implicit in the use of shape grammars is the notion that a painter's style can be completely described by rules. However, painters are capable of producing works that are uncharacteristic of their style, and therefore break some of the "rules" they normally follow. Thus, for example, Hersey and Freedman point out that Palladio occasionally broke the rules specified by Stiny, Gips and Mitchell for the design of his villas [Hersey and Freedman, 1992, p.132]. This does not mean that, in theory, the formal system of shape grammars is incomplete, since on encountering a painting that appears to break a production rule, a new rule may be incorporated to relax the constraints of the system. This, however, leads to two further complications. In the first instance, relaxing the systems constraints increases the chance that the production rules produce a work that is not in the style of the painter. And secondly, and more importantly from a computing perspective, the proliferation of production rules to accommodate outliers increases their complexity and consequently decreases the performance of the respective algorithms. In fact, because shape primitives are usually very simple, the more complex the range of shapes that appears in a given style, the more

[47] Gips and Stiny [1972] outlines the basics of shape grammars. The system is presented more thoroughly in their respective PhD dissertations: Stiny [1975], Gips [1975].

[48] Kirsch et al. [1957] led the team that developed the flatbed scanner.

production rules are required to describe it. Thus shape grammars may be useful for the description of styles that employ a limited set of shapes and apply them repetitively, but the number of rules required to describe a very complex style is prohibitive.

The Second major weakness of shape grammars for our purposes is that the system is descriptive and not analytical. While there are numerous examples of shape grammar based computer algorithms creating art in the style of a particular painter, none have yet been able to analyse any new painting or identify its painter. Kirsch [1997, pp.156-7] acknowledges this limitation and explains,"tools which are most powerful for description are *a fortiori* correspondingly less powerful as analytical devices" and thus suggests that an intermediate solution be found tools that are moderately good at both description and analysis. The root of the problem is that shape grammars are almost entirely descriptive. The process of specifying the formalism is done entirely by a human being, with no help from the computer. Therefore it is not a solution to the problem of classifying new works of art, but to the problem of getting computers to mimic artists. Furthermore, it's demands on the human are not minor. Formal specifications of even simple and repetitive pattern-based styles are rewarded with doctoral degrees, while the task of formally specifying a single figural painter can be a lifelong undertaking. This undermines the rationale behind the present thesis: to facilitate the process of attribution.

Clearly, shape grammars are not ideally suited to the task of computer-aided classification of paintings. The nineties saw a series of more appropriate approaches to the problem devised. These were, in contrast to the high level theory of shape-grammars, practical approaches based on empirical methods, machine learning or statistics. Perhaps the most famous work done in the field is that of Peter-Paul Biro whose work on forensic art history has seen him in numerous newspaper articles and television interviews, and whose services are in much demand by art-houses and dealerships for his scientific attribution and conservation. However, he has not himself published any of his attributions in scholarly journals: his academic articles are all on Rock-art preservation [Michaelsen et al., 2000, Biro et al., 2001]. Biro, on his web-page,[49] describes two important image-processing methods, the more famous of which is fingerprint analysis. The frames and painted surface of the image are scanned for fingerprints that were left during the creative process (such as prints left on the painted surface while the paint was wet, or prints left in varnish before it dried).[50] The other method used by Biro is the recovery

[49]Peter-Paul Biro [2008]: http://www.birofineartrestoration.com.

[50][King, 2000] summarises some of Biro's most interesting cases, including his landmark

of signatures on paintings by band-filtering the area in which a signature is suspected by assuming a particular brush-stroke width, which he used to confirm an attribution of a painting to Eduardo da Rosa.

In contrast with Biro's methodology, which is tailored to individual attributions commissioned by art-dealers or collectors, a small body of scientific literature on more general algorithms leading toward computer-aided attribution has been built over the past 20 years. The Pattern Recognition and Image Processing (PRIP) group at the Technical University of Vienna have been very active in this area launching two separate interdisciplinary projects. The first, unimaginatively titled Art History Project conducted in conjunction with the Austrian art dealer Robert Keil, involved the study of several Austrian miniature portraits, principally using brush-stroke analysis. The second, Cassandra, involves the analysis of infrared reflectograms for brush stroke analysis. Various models for brush stroke segmentation were used. The first involved manual segmentation of the image into strokes by an art historian, from which was developed a model for analysing strokes (called the Model Based Stroke Operator or MBO) based on parameters such as width and orientation of lines [Kropatsch et al., 1995]. In 1998, they trained a multi-layer feed-forward network (MLFN), to develop a stroke detection operator, which performed considerably better than the MBO [Melzer et al., 1998, Sablatnig et al., 1998]. Once brush strokes had been detected, attribution was based on both statistical texture analysis, by means of first and second order statistical measures, and frequency analysis using wavelets.

Textural analysis remains a popular method for digital attribution. Support Vector Machines (SVMs) have been used by Jiang and Huang [2004] to distinguish between two schools of Chinese paintings based on textural features derived from edge-size histograms and autocorrelation, and Li and Wang [2004] has used multiresolution hidden Markov models (MHMMs) to develop a texture model that was used to classify Chinese Ink paintings. Other features have also been proposed for attribution. SVMs have been trained on colour profile curves by Widjaja et al. [2003] to distinguish between Reubens, Michaelangelo, Ingres and Botticelli. More recently there has been controversy surrounding the claim that fractal dimensions could be used to authenticate the works of Jackson Pollock. The original thesis proposed by Taylor et al. [1999b,a, 2007] is that analysis of the drip patterns of authentic Pollocks revealed fractal patterns. However, Jones-Smith and Mathur [2006] have argued that the scale over which the fractal dimension was measured was small enough that both freehand paintings and gaussian

attribution of a Turner in 1993 that sold for $208 000 and a more recent Picasso (no date given).

random motion generate similar fractal patterns. Furthermore they present an example that they themselves drew, dubbed Untitled 5, and show that by Taylor's criteria this work would be authenticated as a Pollock. This controversy serves to highlight the fact that none of the techniques proposed so far can replace a skilled art historian at attribution but that, for the moment at least, they should be used in conjunction with an art historian.

1.4.3 Computer-Aided Documentation of Archaeological Ceramics

Considerable research has been carried out in the fields of computer-aided pottery classification and documentation, although only a handful are directly applicable to the present study. The process of artefact classification can be painstaking. Many artefacts are usually uncovered, many of which will be of little value. Classification of these is a tedious, but necessary process. Furthermore, in the case of painted Attic pottery, the attribution of these artefacts requires very narrow expertise. Therefore, methods that can help automate the process could be of some use to archaeologists. The present study most naturally falls within this discipline although with the distinction that the particular area of archaeological classification is primarily art historical whereas other studies have been concerned largely with shape.

As early as the 1960's statistics were used in the analysis of pottery [Clark, 1962] and computers were used to analyse flint assemblages [Doran and Hodson, 1966]. However, it was only in the late 1970s that computer applications for the analysis and classification of ceramics were first tried. Peter Main's efforts were particularly groundbreaking in this regard. Main [1978] designed a database system for pottery retrieval based on the analysis of their outlines, a topic that would become the subject of his PhD dissertation [Main, 1983]. Shape analysis continued to be the dominant approach to automatic classification. These include the use of beta splines for profile analysis [Hall and Laflin, 1984] and the application of principal components analysis (PCA) and fuzzy models to Chinese Porcelain [Liming et al., 1989]. During the 1990s, Steven Shennon and Paul Lewis developed the GOAD (Graphically Orientated Archaeological Database) project which used the generalised Hough transform (GHT) and other shape descriptors as the basis for an intelligent pottery database [Lewis and Goodson, 1991, Lewis et al., 1993, Durham et al., 1996, 1995].

More recently, shape analysis has been proposed to solve archaeological questions other than typological classification, such as the chronological development of Iron age ceramics at Tel Dor [Ayelet Gilboa and Smilansky, 2004]. In addition there is a considerable literature on the use of shape description and statistical models to analyse fragments or sherds, particularly to determine the type of the vessel and for reconstruction. Maiza and Gaildrat [2005], for example, uses genetic algorithms to develop an implicit measure of distance between sherds and the possible shapes from which they derive. The Pattern Recognition and Image Processing group at the University of Vienna (PRIP) have done considerable work in the classification of sherds for reconstruction [Adler et al., 2001, Kampel et al., 2001, Mara et al., 2002]. Some of their achievements are summed up in [Kampel and Sablatnig, 2007] who describe a system based, in the first instance, on automatic segmentation of Roman burnished ware sherds into rim, base and wall using a model derived from expert knowledge, and the classification of these sherds into one of two types of beaker, two types of pot, three types of plate and three types of jug. A considerable amount of work has also been done by the SHAPE lab at Brown University [Leymarie et al., 2001]. The bibliography of works is enormous and includes, among many other topics, reconstruction from laser scanned profiles and edges of sherds using Bayesian methods [Willis and Cooper, 2002, 2004] and by puzzle matching [Kong, 2002, Aras, 2007, McBride, 2003]. In addition to a number of PhD projects, the SHAPE lab has also developed software for archaeological databases that is used at the great Temple excavation at Petra.

Despite the enormous amount of work that has been done on the image processing of ceramics, there have, as yet, only been two attempts at computer-aided style analysis of the painted surfaces of Attic vases: both are doctoral dissertations [Bishop, 2006, Durham, 1996]. The main rationale behind both these studies is similar to my own: that there are various aspects of archaeological classification and documentation that are painstaking, difficult and occasionally requiring very narrow expertise, that could more easily be undertaken by a computer. The more recent of the two is Bishop's *The Classification of Greek Pottery Shapes and Schools Using Image Retrieval Techniques.* Bishop designed a system that classified vases according to shape (i.e neck amphora, amphora, hydria etc.), and according to what Bishop refers to as “school”, meaning red-figure, black-figure or white-ground. To test the system, a database of 600 images of vases was created: 200 training images and 400 test images. To recognise vase shape, Bishop uses basic topological properties like solidity, extent, eccentricity and some basic proportions and measurements such as area and height. To determine school she used the relative proportions of red white and black in the image of the vase. The basis of Bishop's classifier is a shape retrieval technique based on visual query. A query takes the form of a cropped image of a vase, and the search engine returns the five images that have feature vectors closest to that of the query in terms of Euclidean distance. Despite the simplicity of this technique, Bishop's technique produced excellent results. The closest match belonged to the same group as the query 97% of the time in regard of shape and 100% of the time in regard of school.

More closely related to this study is Durham's *Image Processing and Hypermedia Tools for Archaeological Classification* which arose out of Lewis and Shennan's GOAD project. Durham uses feature spaces derived from the principal components of the generalised Hough transform (GHT) and Visual Moments as the basis for the classification of vases. Durham uses modern Cretan pots to assess the extent to which his classification technique matches human classifications, and then proceeds to test whether the technique works on Attic pots. Specifically, he tests whether the techniques that worked best in the pilot study could distinguish between pots painted by the Antimines painter and those painted by the Swing painter.

For the Cretan pots, Durham uses three classifier designs: k-nearest neighbour (k-nn) and neural networks using visual moments as the feature-space, and hierarchical agglomerative clustering using a distance measure based on the GHT. These classifiers are tested on their ability to distinguish between different pot-types, different pithos types, to identify the village in which the pot was manufactured, and finally to identify the maker of the pot. The results from this study are impressive and are listed in table 1.1

Method	Pot type	Pithos type	village	maker
Moments & k-nn	.98	.97	1	.71
Moments & neural networks	.94	.83	1	.59
GHT & hierarchical agglom.	1	1	1	.85

Table 1.1: Relative accuracies of different classification techniques on Cretan pots

He finds that hierarchical agglomerative clustering works considerably better than the two supervised learning techniques, with neural networks performing particularly poorly. This he attributes to the small sample size, noting particularly that neural networks require around 10 times the number of samples as the number of nodes to perform well.

In the case of Attic pots, sample size was severely limited: 24 samples by the Swing painter and 14 samples by the Antimines painter. Thus, for the study of Attic pots he abandons the supervised learning techniques altogether and uses hierarchical agglomerative clustering exclusively. For this task he measures three different aspects of the pots for comparison: profiles of the pots (vase shape), border decorations (in particular, subsidiary decoration), and human figure details (in particular, bare calves). In respect of vase shape, Durham's system was capable of distinguishing between different types of pots, but not between painters, although it is not certain whether the Swing Painter's *oeuvre* was the work of a single potter.[51] As regards subsidiary decoration, the result was slightly more promising as it grouped the patterns into 5 very homogenous groups. Of the 21 vases tested only 2 were classified to a group that was dominated by examples from the other painter's corpus. The classifier was also moderately successful at matching close-up photographs of the calves of figures drawn by the respective painters. However, these results were not as impressive as the clustering of the subsidiary decorations.

It is difficult to measure the success of Durham's system objectively. Not only does Durham assess each stage qualitatively rather than with a quantitative measure, but the particular classifier that he uses is itself not well suited to any of the established measures of classifier performance. The reason is that the specific clustering algorithm used automatically chooses the most appropriate number of groups into which to classify the dataset. This means that unless the classifier happens to choose the correct number of groups (which happens in the pilot study on Cretan vases but not in the study on Attic vases), one cannot use mean error rate (the proportion of misclassified items) or any similar measures of classifier success. Instead, Durham measures the success of the experiment in terms of how homogeneous each group is. This is, however, done qualitatively without an objective measure of homogeneity specified. Durham's own interpretation of the results is that the sample size is too small to draw firm conclusions but that they are sufficiently positive to recommend further investigation - an assessment with which the author agrees.

Durham's dissertation is an important study of automatic artefact documentation and classification. While previous studies had used automatic classifications on measurable features, Durham tackles the very elusive notion of style, which is difficult to quantify. However, there are a number of improvements that the last 12 years of research into statistical classification techniques and shape recognition algorithms can offer to this genre and the present study aims to extend Durham's work in a number of ways. First of all, the present study has a very different approach to classification: Durham used cluster analysis rather than supervised learning because sample size was too small to use train neural network and because his GHT algorithm did not produce a metric. However in the previous years considerable advances

[51] Böhr [1982] considers the vases he painted to be the work of a single potter, but this is doubted by von Bothmer [1984].

have been made in supervised learning in the small sample environment and these techniques have been used in this present study. The second point of departure is that the present study extends the classification to more than two groups (3 in chapter 3, 5 in chapter 4, and 4 in chapter 5. Finally, the present study demonstrates some practical application of this technique by demonstrating how it can be used in the study of a single painter: the Princeton Painter.

1.5 Outline of Dissertation

This study may roughly be divided into two parts. The first establishes the methodological and theoretical background, while the second is concerned with the actual application of the algorithms to the corpus of the Princeton painter. As already stated in the preface, the next four chapters use a considerable amount of mathematical methods. However, they also provide summaries of the most important elements in italics. The final chapter is more general and most of it does not require specialised knowledge.

1.5.1 Theory and Methodology

Attribution is a classification of art works, or parts thereof, to a particular artist based on a set of examples. This chapter develops a theory and a terminology that allows problems of style and attribution to be expressed in the language of pattern recognition and machine learning. The chapter then proceeds to describe several methods by which these problems may be solved. Particular attention is payed to the fact that the number of examples is generally very small placing severe limitations on the techniques that can be used. A number of strategies are described to overcome this small sample problem. First, principal components analysis (PCA) is used for feature space dimensionality reduction. Secondly, artificial data have been used in other areas to increase the size of the training set and in one of the methods presented (in chapter 3) we use this method. Third, classifiers are combined using ensemble methods to reduce classifier variance. Finally, this chapter explains the methods by which each of the case studies that follow will be assessed, paying particular attention to the difficulties of obtaining reliable performance estimates in the small sample setting.

1.5.2 Morellian Analysis: The Princeton Painter's Knees

This chapter introduces a novel approach to automated attribution by applying shape descriptors to implement the technique of Giovanni Morelli. As has been explained in 1.1.3 Morelli believed that the clearest indicator of the artist who rendered a painting is the manner in which he rendered minor details, particularly anatomy such as ears and hands. The methodology he employed, although widely criticised in traditional art historical circles, is the primary methodology employed by archaeologists specialising in the attribution of Attic pottery. Thus, this chapter is concerned primarily with getting a computer to conduct style analysis in the same way that Humans have done in the past. Since Morelli's technique involves comparing neatly defined units of a painting: hands, ears etc., the shape recognition algorithms are well suited to distinguishing between the basic forms associated with particular artists. In particular using some well established shape descriptors, LDA, QDA and nearest neighbour are capable of classifying between knee types associated with the Princeton Painter, Exekias and Group E when trained on each of the feature spaces. Moreover, it appears as though the classifiers still perform relatively well even when trained exclusively on artificial examples drawn by an art-historian in the manner of the painters. Finally, a majority vote ensemble of all the classifiers used in the study far outperforms any of the individual classifiers alone.

1.5.3 Proportions: Male Human Heads

While the previous chapter was concerned primarily with the adaptation of traditional methods of style analysis for computer automation, the present chapter deals with an entire new method of analysis that can only be conducted by a computer. Human male figures were the primary subjects of Attic black figure scenes, and the expressive elements of the painting are usually captured by his gestures rather than his facial features (which are generally expressionless). Because the facial expressions are invariant to the subject of the paintings (and therefore do not change much within the corpus of a particular painter) they are potential candidates for attribution. Using the relationships between the positions of various major facial features (eyes, mouth, nose, forehead) on bare male heads, classifiers achieve reasonable accuracy when distinguishing between the heads of 5 different black-figure artists and 100% accuracy distinguishing these from real human heads.

1.5.4 Vase Construction

It has long been established that the shape of pots is a useful indicator of the internal chronology of a painter, as illustrated by Technau [1936] and Mackay [1981] on the corpus of Exekias. Hansjörg Bloesch [1940] showed that vases tended towards slenderness towards the end of the sixth century BCE. Furthermore, scholars like von Bothmer have shown that an examination of the shapes of vases may be used to attribute vases to a particular potter. This chapter investigates techniques for pottery classification according to style and applies these findings to the work of the Princeton Painter to answer the long-standing question of whether the Princeton Painter worked together with a single potter. The question has serious ramifications regarding the internal chronology of the Princeton Painter.

1.5.5 Conclusion and Postscript

The final chapter deals with two issues regarding the implementation of the methods in this study. First, this dissertation has concentrated on the classification of minor details on vases according to the painter that produced them. However the attribution of an entire work is generally made after consideration of many minor details of that work. Usually it is up to the intuition of the expert as to whether an attribution should be accepted or not. There are, however, a number of scientific methods that may be used to combine the results of attributions, and although rigorously proving these is an immense project and beyond the scope of this dissertation, we present a number of methods that may be fruitful areas for future research. Secondly, for largely historical reasons, the discipline of Classical Archaeology employs black-and white photographs as the standard medium by which images of vases are made available to scholars. Currently most of these photographs may be obtained online through the Beazley archive. Unfortunately, these are insufficient for many attribution tasks since they distort various features, and obscure others. Furthermore, many of the photographs are of poor quality, and in many cases, only photographs of one side a vase exist. This chapter explores the potential for alternative methods of publishing images of artefacts and suggests that if these were adopted as a standard by the Corpus Vasorum Antiquorum, it could ameliorate some of the difficulties encountered in this dissertation and furthermore will facilitate further research in the field.

Chapter 2
Theory and Methodology

This chapter presents the core theoretical and methodological concepts that are used in the dissertation and is therefore essential reading even for the reader who is only interested in the results presented in even one of the subsequent chapters. Much of the theory assumes a background in linear algebra, probability theory and statistics. In particular, the chapter assumes an understanding of joint, marginal, and conditional probabilities, as well as expected values and covariance. However, for readers with a basic mathematical background the conclusion summarises the most important issues and gives a summary of the main terms required to understand the findings

Many of the concepts in this chapter are presented with their derivations, but these are not necessary for understanding how these concepts relate to this study. Instead, the relevant discussions are more important. For the reader who is familiar with pattern recognition, much of this chapter may be very familiar. Nevertheless, the first section of the chapter is still necessary because it both defines the terminology used in the subsequent chapters and because it explains attribution as a pattern classification problem. In addition, two other sections are important for understanding the rest of the dissertation: PCA90/70 (section 2.3.2) and the methods used in this dissertation to report results of experiments (2.5). For the reader who does not wish to wade through the technical details at all, there is a brief summary at the end of the chapter.

2.1 Introduction

The purpose of this chapter is to develop the key theoretical concepts on which the dissertation is based, to define the technical vocabulary used, and to explain the methodological basis for the experiments undertaken in chapters three, four and five. The first section defines a small technical vocabulary that allows questions of style to be phrased in statistical terms and allows questions of attribution to be posed as pattern classification problems. The second section defines the classification task and explains the methodology of pattern recognition, paying particular attention to the algorithms and strategies that are common to the rest of this dissertation. The topics covered are a decision rule construction, feature extraction, sampling, parameter selection, evaluation and ensemble methods. Of particular importance throughout these explanations is the issue of how standard concepts of pattern recognition methodology need to be adjusted in the small sample setting. Some of the concepts in this chapter are dealt with formally with mathematical proofs and others are treated philosophically. This is to avoid duplication since the chapters that follow use slightly different methods, and if a concept is relevant only to one chapter, its rigorous treatment is differed to that chapter. On the other hand if the concept is important to the whole thesis, it is treated rigorously in this chapter

2.1.1 Pattern Classification Terminology

Informally, one may define pattern classification as the task of assigning objects to a set of classes based on some features of these objects. Three examples of pattern recognition systems should give some idea of the range of algorithms that can be described as such. Example 1 is a machine that tests for a disease, say diabetes, based on a measurement of a patient's fasting plasma-glucose level in mmol/l. In this example, the objects to be classified are the patients, there is a single feature - the blood-glucose level, and the classes could be negative and positive. Example 2 is a machine that calculates the risk that a loan applicant will default. In this case, the objects are the loan applicants, the features are a set of statistics derived from the applicant's financial history, and the classes may be a set of categories like high, intermediate, low and very low. Example 3 is a little more complicated: a machine that takes as the set of features the results of a loan applicant's default risk assessment and the expected interest (in the lowest unit of currency, say cents) made on such an applicant if he or she were not to default on the loan, and on this basis makes a decision as to whether to grant or to decline such a loan. In this final example, the possible classes are simply decline or accept. Example 1 and 3 are a special class of algorithm with binary output since they only classify the objects into one of two classes. Such simplicity greatly simplifies many of the issues in pattern classification, but the examples in this dissertation are multi-class.

Now that pattern classification has been described by example, a more formal definition is provided. Since the field of pattern recognition has diverse origins, there is no single established terminology and terms are often defined by the author. This study uses the following formalism to describe pattern classification problems. Assuming a finite discrete set of C class labels $\Omega = \{\omega_j; j = 1 \ldots C \in \mathbb{N}^+\}$, and a set of all classifiable objects $\mathcal{U}$, the problem of pattern classification is to design a function that assigns a unique class label in Ω to every object in $\mathcal{U}$. The assignment is based on a set of d measurements that can be taken of the objects in $\mathcal{U}$, which may be represented by a set of variables or features, $X_1 \ldots X_d$;$d \in \mathbb{N}^+$. The n dimensional vector space spanned by all possible values of these variables is called the **feature space** and represented by $\mathcal{F}$. If a complete set of measurements is taken of an object, it may be represented by a vector $\mathbf{x} = [x_1 \ldots x_n]$, called a **feature vector** or a pattern. In this way, each object may be represented as a point in $\mathcal{F}$. The pattern classification problem may be solved by a mapping $f : \mathcal{F} \rightarrow \Omega$, called a **decision rule** or **classification rule**. An algorithm used to construct and implement such a rule is a special kind of learning algorithm which in this study will be called a **classifier**. If the classifier constructs the decision rule using a **design set** $\mathcal{D}$ composed of n pairs of correctly labelled feature vectors then we refer to the classifier as a **supervised classifier**. This type of classification may be realised by a mapping $f : \mathcal{F} \times \mathcal{D} \rightarrow \Omega$. Although it is normally assumed that $\mathcal{F}$ is a subspace of $\mathbb{C}^d$ or $\mathbb{R}^d$, this is not always true since feature spaces may take on a variety of different mathematical forms. In this study a single item within a test, design or validation set will be referred to as a sample.

These formalisms may be illustrated on the three examples. In example 1, $\Omega = \{\omega_0 = \text{negative and } \omega_1 = \text{positive}\}$ and $\mathcal{F} = \mathbb{R}$ since plasma-glucose level may be represented by a real number. A possible decision rule could be $\omega_n, n = [x > 7.0]$, where x is the feature vector associated with the patient being diagnosed and [..] represents a casting of a logical expression into binary such that 0 is false and 1 is true. Informally this translates to diagnose positive (ω_1) if plasma-glucose is more than 7 and negative otherwise (7 mmol/l is the typical threshold for diagnosis of diabetes). In example 2, $\Omega = \{\text{high, intermediate, low, very low}\}$ and $\mathcal{F}$ is the relevant applicant's financial statistics. In example 3, $\Omega = \{\text{grant, do not grant}\}$, and $\mathcal{F} = \{\mathbb{Z}, \{\text{high, intermediate, low, very low}\}\}$. The feature space of example 3 is not simply a subspace of $\mathbb{Z}^d$ because it is composed of both a natural number variable and a discrete variable. For example 3 a possible (though sub-optimal) way to construct a decision rule could be as follows. Assign the following values to the applicants risk: 1 to very low, 0.8 to low, 0.1 to intermediate and 0.01 to high. Then multiply the risk by the potential reward and grant the loan if this value is above the amount of the loan modified to take into account uncertainty and the lending rate. Formally this could be written as $\omega_n, n = [r * p > m]$ where r is the value assigned to the risk level, p is the expected profit and m is the modified loan value.

Example 2 may further be used to illustrate supervised classification. Imagine that the categories of high, intermediate, low and very low risk were defined by induction over the bank's existing database of the financial details of clients who've taken loans in the past. A group of accountants samples a large number of clients and by analysing the data, group the clients into a set of 5 classes based on what appear to the accountants to be salient variables derived from the clients' financial histories. The overall default rate of clients within each class determines the label applied to each member of that class. These are used as the class labels for the classier. Then a database is constructed such that each client is described only by the salient variables and the class to which he was assigned (high, intermediate, low, very low). In other words, the database consists of a number of ordered pairs $[\mathbf{x}, \omega]$ where $\mathbf{x}$ is a feature vector representing the variables defined on the client's financial history, and ω represents the class to which he/she belongs. An algorithm is devised to induce a rule based on some structure in the data that unites members of each class and separates them from members of other classes (bear in mind that some of the work has already been conducted by the accountants who selected the salient variables). The induction rule can then be applied to any new client who will duly be classed according to the class to which they are most similar. This is an example of a supervised learning algorithm. The salient variables span the feature space $\mathcal{F}$, and the database is the training or the design set $\mathcal{D}$.

One particular method of generating a decision rule is to partition the d dimensional feature space by a (typically $d-1$ dimensional) surface, or a number of surfaces, whereby each partition has associated with it a class-label, and class membership is determined by which partition an object's feature vector resides. Figure (2.1) shows two simple classifiers of patterns in a two-dimensional feature space. The true classes of the objects are indicated with dots and crosses respectively, while the predicted classes are represented by the class labels $\{\omega_1, \omega_2\}$. The decision rules in each case have been constructed by partitioning the feature space into two sections using a function. In (a) the function is a curve and in (b) it is a line. If, as in the two examples above, a classifier constructs a $d-1$ dimensional surface in the feature space and assigns class labels based on which side of this surface they lie, then the surface is called a **decision surface**. In supervised classification, the parameters of the surface are determined to best fit the training data, and new items may be classified according to the partition in which their feature vectors reside.

A popular and important method of designing a decision rule is through a **discriminant function**. This is a mapping from the feature space onto a scalar variable where classification is based on the value of the variable (for example classify as ω_1 if $f(\mathbf{x}) > 1$ else ω_2). Both example 1 and example 3 are discriminant functions - 1 because it only has a single scalar variable as the feature space, and 3 because the final decision is made ultimately on the basis of a single scalar product. In cases where there are more than 2 classes, multiple discriminant functions may be used and classification is to the class label that maximises the discriminant function. LDA (2.4.3), a frequently used classifier in this dissertation, is based on a discriminant function.

With the broad framework of pattern classification and the basic terminology associated with decision rule construction defined, an important and often overlooked issue is now addressed. That is, how does one construct a

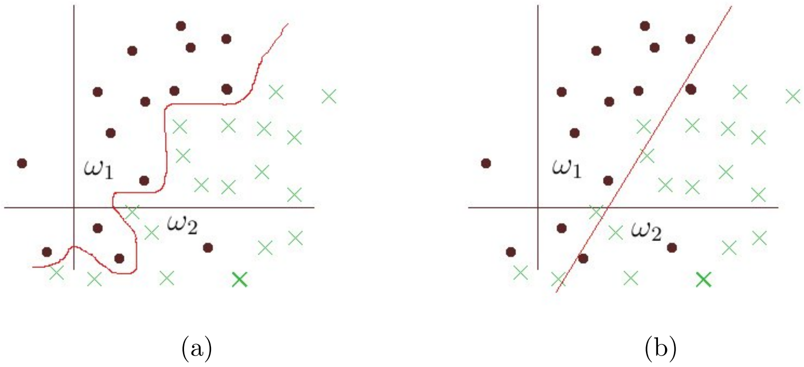


Figure 2.1: Two feature spaces partitioned by decision surfaces. (a) is a free curve and (b) is linear.

feature space or, more practically, how does one convert objects into feature vectors. There are a variety of terms used in the literature and terms are not used necessarily unambiguously across sub-disciplines. For this reason, and to meet the needs of this particular study, the relevant terms will be defined in a way that may be unfamiliar to readers who are already well-versed in machine learning and pattern classification. It is intuitively clear that objects in the real world need to be described in a mathematically meaningful way before inferences about their nature or behaviour can be made using mathematical principles. In the case of pattern classification, it is typical to describe certain attributes of the object and encode these attributes in a feature vector, as has already been described. This study distinguishes between two types of algorithms that have feature vectors as their output, or in other words that are mappings into feature spaces: primary and secondary **feature extractors** which are defined below.

Formally we may define primary and secondary feature extractors as follows. A **primary feature extractor** is a mapping $f : \mathcal{U} \to \mathcal{F}$ and the a **secondary feature extractor** to a mapping $f : \mathcal{F} \to \bar{\mathcal{F}}$. The distinction is subtle but important in the context of this study. A primary feature extractor is any algorithm that takes measurements of an object and outputs a feature vector. For the sake of this study, objects will include digital images. Informally, a primary feature extractor describes real-world objects in a way that allows them to be represented by feature vectors, and hence in such a way that they may be subject to pattern classification. In addition, however, it is sometimes useful to transform one feature space into another so that certain properties which are not easily distinguishable in the first space are distinguishable in the second. This is a secondary feature extractor which takes as its input a feature vector and outputs another transformed feature vector. A simply example that is common in many scientific fields is the transformation of an audio signal from an amplitude-time signature into a power-frequency spectrum. Both of these, provided they are discrete, can be represented by feature vectors. However, if one is interested in classifying the audio signal into different possible musical notes, then the power-frequency spectrum will be a much easier feature space to work with than the amplitude-time space. In the context of art, primary and secondary feature extractors are defined again in 2.1.2). Feature extractors can be chained together to produce different feature spaces. For interest sake, it is worth pointing out that an algorithm that applies a primary feature extractor to an object and then applies a secondary feature extractor to the output of this object is itself also a primary feature extractor. A commonly used secondary extractor in this dissertation is principal components analysis (PCA explained in 2.3.2), although a variety of different extractors are used in chapters 4 and 5.

2.1.2 Attribution as Pattern Classification

Attribution is the act of classifying a work of art (or fragment thereof, or a group of works or fragments thereof) according to who produced it.[1] Traditionally, there are many methods used to attribute a work of art. Some of the most common methods include: a study of the history of the work, a physical analysis of the medium on or of which it is made/rendered, and a close examination of the style in which the work is rendered. A detailed history of the artifact could be established by scrutinising the available documentary and literary evidence pertaining to the work, such as the collections in which the artifact was housed, where it was found, which scholars have examined it, what their special skills were, and what their findings were. A scientific examination of the physical medium may reveal important clues such as where the artifact came from, how old the medium is, and on occasion even reveal clues such as fingerprints. Finally, an analysis of the style provides possibly the most obvious clues - those details of the technique that are unique to the painter. In the case of Attic vase painting, this has often been the only clue, since the history of the vase is lost, we have no documentary or literary evidence of any known painter, and scientific procedures are not accurate enough to provide useful information about the date of production.[2]

[1] "Who" could refer to a number of different things including an artist, a group of artists or even a stage in an artist's career.

[2] Thermoluminescence, for example, is only accurate to a few hundred years and therefore only useful for authentication.

Figure 2.2: A Princeton Painter Ear rendered with an 'S' shaped incision.

By the informal definition of the pattern classification task described earlier, it is clear that attribution may be defined as a pattern classification task. In this section a formal terminology is defined that allows questions of attribution to be posed as problems of pattern classification. The concepts and terms developed here will be illustrated using a hypothetical attribution problem: The Princeton Painter regularly renders the ears of his male figures with an "S" shaped incision (see figure 2.2).[3] The hypothetical attribution task is to distinguish examples of ears as being either "S" shaped or non-"S" shaped. The most crucial terms relate to the observable elements of style. Morelli referred to *grundformen* which were manifestations of a painter's method, such as the peculiar way in which they rendered hands or heads. For the purposes of this theory, a generalisation of this concept will form the foundation of the theory. So as not to exclude any feature that may be of interest to any researcher, we use a very broad notion that may be narrowed down according to the application.

Definition 2.1.1. *a **stylistic form** is a manifestation of a work of art or a collection of works of art, either hypothetical or real, subject to the restriction that it may be described numerically. A set of forms that all meet some specific criterion is called a **form set**.*

This definition is broad enough to represent any measurable quality or a

[3]Detail from the obverse of London B212 **R9**.

physical measurement on a work of art. For example, an ear on a painting is a form as it may be represented numerically by a digital image, as can the set of all ears in a painter's corpus. The numerical description does not refer only to spacial dimensions, but any numerical feature. For example, quite abstract properties of a painter's work may be described as forms, such as neatness, provided that some numerical method of describing neatness is available. Of course, the more interesting possibilities are those in which the abstract quality that is being measured has no name but is defined purely by the mathematical formalism of measuring it. For example, in 4.3.3 the relative angles between major facial features of the male figures are used to distinguish heads rendered by the Princeton Painter from those rendered by his contemporaries. The quality that is measured is very abstract indeed, but it is nevertheless still a form.

The convention employed is to use γ with an appropriate subscript to denote a form and Γ with an appropriate subscript to refer to a form set. For our example we may define two sets, Γ_S and $\Gamma_{!S}$ which are respectively the set of all male ears that are "S" shaped and all those that are not. If we refer to the collection of all sets of male figure ears for which no attribution has yet been made as Γ_{ears} then the attribution task, in this case, will be implemented by a classifier that assigns to all the elements of Γ_{ears} a class label indicating membership of either Γ_s or $\Gamma_{!s}$.

The constraint that a form must be numerically describable implies the existence of a sensory apparatus and suitable algorithms to make these descriptions and an abstract mathematical space in which these descriptions may be represented. Here we employ the terminology developed in the preceding section of a primary feature extractor that converts forms into numbers and a feature space in which these forms may be represented and manipulated. More formally

Definition 2.1.2. *A* ***primary feature extractor*** *defined on a form set* Γ *is any function that assigns to every form* $\gamma_n \in \Gamma$ *a vector* $\mathbf{x_n}$ *describing it called the feature vector of* γ_n *The vector space spanned by the feature vectors of all forms in* Γ *is called the feature space of* γ *and is usually represented by* $\mathcal{F}$.

Thus, in the case of the "S" shaped ears, the process of photographing, scanning, image processing, segmentation, cropping and the application of a shape description algorithm to the final image, is all described by the primary feature extraction function. An example of the primary feature extraction process is the scanning of ears, the extraction of the skeleton, and the application of a 100 element vertical projection descriptor to the image to describe its shape as a 100 dimensional feature vector.

We may also define transformations from one feature space to another.

Definition 2.1.3. *A* ***secondary feature extractor*** *is any mapping* $f : \mathcal{F} \rightarrow \bar{\mathcal{F}}$ *where* $\mathcal{F}, \bar{\mathcal{F}}$ *are different feature-spaces.*

A trivial example is applying a log function to feature vectors to reduce the effect of outliers. An important class of feature extractors transform data from a high dimensional feature space into a considerably lower dimensional feature space without losing much of the information. For example, Principal Components Analysis (PCA) (2.3.2) may be used to find a subspace which describes most of the variation in the data set, but using far fewer variables.

A physical object or an abstract quality in itself cannot be used as an input to a classifier, but measurements taken on the object can. Thus the primary feature extractor is the 'bridge' between the domain of real objects and the mathematical constructs that allow them to be analysed. The reason for making the distinction between primary and secondary feature extractors is that this study often uses both the initial measurements for one task and a transformation of these measurements in another.

In the process of attribution, the question of "who" produces the forms themselves is quite pertinent. It is naïve to assume that this is always a painter. For example, one may want to be more specific and identify both the painter and the period in the painter's career. Alternatively one may be interested in assigning a work to a school or a genre. In fact, some archaeologists may be happy to distinguish between two material cultures. Thus, attribution is not simply the classification of an object (or a set thereof) to a painter. A second issue is that, in the case of the artistic personalities described by Beazley, we cannot be sure that we are indeed attributing to a single artist at all. As stated in the introduction the aim of this study, in any event, is not to recognise the work of real artists, but rather the personalities identified by Beazley. Much of the terminology used by Beazley and the pioneers of the field were borrowed from art history and reflect concepts that are misleading in the context of archaeology. For example, Beazley uses terms like "manner of", "near" and "workshop of" to refer paintings and vases that are similar in style to a particular artist, but not close enough for a secure attribution. Such terminology implies that vase painters worked in workshops and schools in the way that artists in Renaissance Europe did, but we do not know that this was the case. Therefore a more general term is defined here which not only avoids the misconceptions spelt out above, but also has some appealing properties.

Definition 2.1.4. *If we denote the set of all possible forms as* Γ*, then a* ***stylistic agent*** $\pi(\theta, \Gamma)$ *is a probability distribution over* Γ *relying on some parameter vector* θ*.*

In other words, a stylistic agent is a process that produces forms with a fixed probability. For example, given two agents: the personality Beazley called "the Princeton Painter" and "Exekias", denoted by π_1 and π_2, we may be interested in the attribution of an ear. The ear has an "S" shape, and such ears appear with some regularity in the Princeton Painter's corpus, but not in the corpus of Exekias. From the perspective of the pattern recognition problem, all that matters is the respective likelihoods that an ear of Exekias and the Princeton Painter would take the form of an "S". Thus the only information required is the probability distribution over the set of the forms.

2.1.3 Formally Stating the Attribution Problem

Informally, using the terminology developed thus far, we define the attribution task as the classification of a form based on its feature vector to the agent that produced it. More formally, given a set of agents $\Pi = \{\pi_c : c = 1 \ldots C\}$ and given a set $\Omega = \{\omega_c : c = 1 \ldots C\}$ where ω_c represents the event that the form under consideration was produced by π_c, then we may define the attribution task as determining which element of Ω represents the true state of nature. Solving this problem then may be done in a number of ways. For example, if the artist is alive, interviewing him and understanding his personality would probably provide some insight into the way he worked. This information might help us understand the shape of the distribution that governed his production of forms (even though this information would be very difficult to quantify). An area of enquiry that has recently become active is the neuroscience of art (such as Ramachandran and Hirstein [1999]) which may eventually allow us to understand the processes that underlie the creative process, and perhaps allow us to parameterise this process.

However, neither of these avenues are open to scholars working on Attic pottery at the moment, so we propose an alternative method - using data collected from their extant works together with art historical intuition to design classifiers to estimate the distribution, and to test these estimates empirically. Unlike many other classification tasks, such as medical diagnosis or stock portfolio risk analysis, we may safely assume that the costs of misclassification are equal. This assumption is well justified in most cases, since when one mis classifies a painting, it rarely matters whether this misclassification is to the Princeton Painter or Exekias.

A typical attribution problem of classifying a painting to a painter will usually be accomplished through an examination of some of the smaller details of the painting. The choice of which details to study will be motivated largely by art historical theory or intuition. Thus, one may view the problem of attributing a painting as a combination of smaller classification tasks which must be combined to obtain the final classification. Methods by which this can be achieved are discussed in 6.2.2. But by the definitions described above there is no conceptual difference between the overall classification and the smaller classifications from which it is built, since they are all examples of classifying forms to agents. Of course, since the form is completely described by its feature vector, we may consider the attribution problem from the perspective of assigning the class labels on the basis of the feature vectors.

2.1.4 Supervised Classifier design

There is no single established flow-chart that describes all possible classifier designs. In fact, this study employs three different classifier designs in the three following chapters. However, common to all of these, and common to most classifier designs are the following stages:

1. Data collection
2. Feature Selection and Extraction
3. Validation and Model Selection
4. Training
5. Assessment

These stages will be discussed in turn although the order in which they are discussed will be presented thematically to allow proper development of the most important concepts. Briefly, data collection is the process of selecting the objects that will be used to design and test the classifier. Feature selection and extraction is the process of measuring the objects and selecting feature spaces that are suitable for the purposes of this classification. Model selection, validation, and training are often repeated in turn. Training is the process by which a decision rule is constructed, while validation tunes the parameters so as to maximise the performance of the classifier and model selection uses some criterion to select between different models for the classifier design (validation is a form of model selection). This often involves training and evaluating the classifier with various different parameters and selecting the one with the best performance. Finally, assessment is the process of estimating how well the classifier will work when applied to new data (i.e. data not used in the process of training or validation.)

While there is considerable variation in the order in which these steps are carried out, and the number of times each step is carried out, these stages are represented in some form in most classifier designs. However, the exact boundaries of each stage are blurred. For example, the boundary between feature extraction and assignment is quite blurred since many decision rules implicitly map the feature space into another space in which the assignment is carried out.

2.2 Data Collection

Before a classifier can be designed, particular in the case of supervised classification, a set of real-world objects is required. The process of selecting these objects is the subject of this section. Of some importance is the fact that for this kind of research, since the available objects are determined by survival and availability, serendipity plays a large role in the collection of data and it is important to consider the ramifications of this (2.2.2).

2.2.1 Test, Validation and Design Sets

Once the data has been collected it is often divided into three different sets: the **design set**, the **validation set** and the **test set**. The purpose of the design set is to train the classifier so that it learns how to distinguish between members of the different classes. The way in which the classifier learns depends on the type of algorithm that is used. Some classifiers have certain parameters that may be fine-tuned, but which require a set that is independent of the design set. The logic is that once trained, further tuning may be done to find optimal parameters by testing a range of parameters on the validation set and selecting the parameter that maximises some performance criterion. Finally, the test set is used to estimate the performance of the classifier. The method of keeping a separate set for testing is called the holdout method. Since classifiers are not typically designed to analyse a particular data set (a host of more suitable multivariate statistical techniques are more suited) but to aid in decisions concerning new and unseen data, its performance is best measured on data that the classifier has not 'seen' during the learning and validation phases. Particularly important is that testing the performance of the classifier on data that it has already seen leads to an optimistically biased estimate.

The problem with dividing the data into three different sets is that the data collected for any classification task is finite and thus a large design set implies a smaller validation and test set. A small design set often means a less expert classifier, while a small test set means a less accurate estimate of the classifier's performance. Thus a number of work arounds have been developed. Resampling techniques, for example, allow design data to be used for tuning and even evaluation. Often this involves estimating the bias involved in using the design set for both training and evaluation (see 2.5.2) and then compensating for this bias.

2.2.2 Sampling

Most classifier designs assume that the data are independent and identically distributed. That is, each of the samples in the data are drawn from the same distribution and the probability of each sample being drawn is independent of any other sample having been chosen. This means, in essence, that the sampling procedure should be random. This also, importantly, implies that that the test and design sets are sampled from the same distribution. Failure to satisfy the i.i.d criteria will bias the data and the consequences of biased data requires some discussion, particularly in the context of this study in which these criteria are not always met, and it is important to consider how this affects the interpretation of the results. There are two concerns which are of importance, how the bias effects the true performance of the classifier and how the bias effects the estimates of the true performance. Which of these is worse depends on the nature of the study. For the purposes of this study the most serious consideration is if the performance estimate is optimistically biased since this is a proof of concept and optimistic performance estimates might misleadingly suggest a concept has been proven when it hasn't. On the other hand sampling bias that negatively impacts on the actual performance of the classifier and a pessimistically biased estimate are similar since they both report results that suggest the classifier performs poorer than it actually does. In the case of a pilot study, the danger is that results that appear to be poor may result in a good classifier being rejected. On the other hand, if it is not rejected, larger scale studies with better data may result in improved performance.

There are however, cases where sampling bias might improve the performance of the classifier and not affect the accuracy of the performance estimate. This is the case when artificial data are used to increase the size of the design set. It seems intuitively likely that induction is improved with more data (discussed in more detail in 2.3.1). Therefore, increasing the size of the design set with artificial data that are not sampled from exactly the same distribution as the test set (violating the identically distributed assumption) may improve performance if the artificial samples are close enough that the increase in the size of design set more than compensates for any bias that may be introduced. In such cases, violating the i.i.d. assumption is beneficial and such a system is used in chapter 3 where forms drawn by the art historian in the manner of a particular painter are used to increase the size of the design set, and it is shown that decent classifiers may be built using the artificial data alone as the design set. Since the test set is still made up of the real world examples, the performance estimate may still be trusted.

When the test set and the design set are both not sampled i.i.d. (violating the independence assumption) the consequences may be more severe than if the test set and design set are sampled independently from slightly different distributions. The primary danger in this situation is that the sampling bias may not only affect the accuracy of the performance estimate, but also obscure the interpretation of the results. In the case of this study, there is an obvious bias, in that survival of Attic vases is not entirely random. There are a number of factors that may contribute to a vase's survival. First, the range in the quality of ceramic ware was great. Some vases were appreciated so much that they were buried with the owner (and in some cases even used as the urn) while others were dumped and used as landfill. If quality vases were more likely to be buried with the owner and poorer examples more likely to be used as landfill then this would bias our sample toward vases of higher quality. This is because the sherds of a vase found in a grave are likely to be very close to each other, whereas this is not the case when vases are disposed of in the open and get moved about by erosion. This means that it is easier to attribute a vase found in a grave than one thrown away. Thus it may be that the vases that are already attributed (from which our sample is taken) may well be the better examples. The second bias is as a result of vases being exported. Only a small percentage of the vases extant today were discovered in Athens - the vast majority were discovered in Italy, suggesting they were exported. Again, it may well be that the exported works represent the better examples. Thus, our data for each painter may well be biased toward the better works in his *oeuvre*. Because these works compose the design set, any classifiers designed in this way are likely to be better at recognising the better works in the painter's corpus than the worse ones. Since we don't know the degree to which the various factors bias our classifier, and we don't know the ratio between poor works and good works of any painter, there are limits to the degree to which we can estimate the true accuracy of our classifier.

However, the goal of this dissertation is not to design a classifier that recognises different painters, but rather to prove the principal that a computer may be taught to attribute in the same way that an art historian can. In this case, the samples are exactly the same samples that Beazley used, and we may assume that they are each independently chosen. The assumption and the hope is that new samples will be classified in more or less the same way that Beazley would have attributed them.

2.3 The Feature Space

Choosing what features to use for the classification task is an important step. It is important to realise that the human designer plays an important role in this process by making the first decision on what features are salient or diagnostic, although they may use automatic processes to actually make the measurements. Once these initial measurements have been taken, they may be transformed in a number of ways, and there exist a number of automatic computer techniques to further refine the feature space for classification. One of these is Principal Components Analysis (PCA) which is widely used in this dissertation, and will be described here (2.3.2). Others are specific to the task at hand and are instead described in the relevant chapters.

There are a number of consideration when choosing a feature space. First, the feature space needs to be meaningful in respect of the classification task. In other words, if the data on which a classifier is trained are meaningless to the task, then we should not expect it to perform above chance. A trivial example illustrates this. If the classification task is to distinguish between people based on their risk of mortality over the next year (say into high, medium and low risk), then it is intuitively obvious that regardless of the learning algorithm used, the features 'eye-colour' and 'height' would be a much poorer basis for classification than 'age' and 'low density cholesterol levels'. A second important issue is how easy it is to take the measurements. The process of taking measurements can be quite complicated and expensive, as it involves a physical apparatus as well as suitable algorithms to transform the raw data from the physical apparatus into a meaningful feature space. A third and very important consideration is the number of variables required to describe each object. If we assume that each variable is a real number then $\mathbb{R}^n$ is the feature space and the number of variables is n, the dimensionality of the feature space. Regardless of the classifier used, adding variables does not guarantee better discriminating power, and in particular, when the number of design samples is small by comparison with the dimensionality of the feature space, then a number of difficulties arise. This is the small sample problem and is significant in this study where sample size is limited.

2.3.1 The Small Sample Problem

In many classification tasks, the performance of a classifier improves as the number of training samples increases, but the relationship between performance and the number of variables is more complicated. Empirical evidence suggests that for a given classifier trained on a fixed size sample, as the number of variables increases, performance increases to a point and then starts to deteriorate [Braga-Neto, 2007]. This is illustrated by figure 2.3 which shows the performance of a classifier versus the number of variables. Thus, while taking more and more measurements might increase the amount of salient information available to the classifier, there are other factors that may reduce the performance if the number of samples isn't also increased.

Small sample problems are not limited to performance, but almost every aspect of classifier design needs to take sample size into account. In attribution studies, the small sample problem will probably always be a factor since artists (especially dead ones) cannot be expected to produce more examples of their technique if more design data are needed. In this study, the small sample issue is so important that it will be discussed in theory here, but in the relevant sections of this and subsequent chapters, the practical implications and the measures used to ameliorate the problem will be discussed in detail. Some of these practical issues are: some classifiers, like LDA (2.4.3), simply can not work when the number of training samples is smaller than the number of variables; when samples are small, one cannot afford to partition the data into training, validation and test sets (2.5); and when the number of test samples is small, the variance of the error estimate is large (2.5.1)

2.3.1.1 The bias-variance decomposition

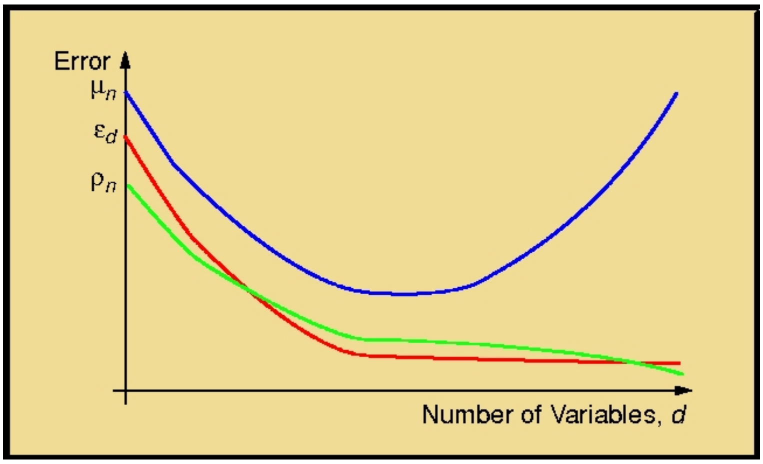


Figure 2.3: Performance as a function of the number of variables. From [Braga-Neto, 2007, p.4]. Here ε_d is the lowest achievable error, μ_n is the true error, and ρ_n is the error on the design set. As the number of variables increases, the training error decreases, while the true error (i.e. on novel instances) reaches a minimum and then increases.

One way of understanding the small sample problem is in terms of bias and variance. This is most easily understood in regression where the attempt is, given a design set $\mathcal{D}$ and an unknown real-valued function f, to find an estimate of the function $\hat{f}$. The square error of the estimate is defined as the squared difference between the estimate and the true (and unknown) function. If we estimate the function based on some learning algorithm then the mean square error of that estimate may be defined as the expected squared difference over all training sets, or formally MSE $= E_D(f - \hat{f})^2$. For the rest of the derivation, the abbreviation E will be used for E_D. Adding the term $\mathrm{E}(\hat{f}) - E(\hat{f})$ allows the following decomposition of the error.

$$MSE = E(\hat{f} - f)^2 \tag{2.1}$$
$$= E[\hat{f} - E(\hat{f}) + E(\hat{f}) - f]^2 \tag{2.2}$$
$$= E[(\hat{f} - E(\hat{f}))^2 + (E(\hat{f}) - f)^2 - 2(\hat{f} - E(\hat{f}))(E(\hat{f}) - f)] \tag{2.3}$$
$$= E[(\hat{f} - E(\hat{f}))^2] + E(E(\hat{f}) - f)^2 - 2E[(\hat{f} - E(\hat{f}))(E(\hat{f}) - f)] \tag{2.4}$$
$$= VAR^2 + \beta^2 - 2E[(\hat{f} - E(\hat{f}))(E(\hat{f}) - f)] \tag{2.5}$$

The term after $VAR^2 + \beta^2$ can be shown to equal zero as follows:

$$2E(\hat{f} - E(\hat{f}))(E(\hat{f}) - f) = 2E[(\hat{f}E(\hat{f})) + E(fE(\hat{f})) - E(\hat{f})E(\hat{f}) - E(f\hat{f})] \tag{2.6}$$
$$= 2[E(\hat{f})^2 + fE(\hat{f}) - E(\hat{f})^2 - fE(\hat{f})] \tag{2.7}$$
$$= 0 \tag{2.8}$$

Thus the MSE of the estimate can be decomposed into a square variance term and a square bias term. The bias term is invariant under changes to the design set, while the variance is dependent on it.

Classification is conceptually very similar to regression, except that instead of a real valued output, a classifier has a categorial output. Therefore intuitively it would seem likely that a similar bias and variance decomposition should be possible for classifiers. However, there is little agreement over what is meant by bias, variance or mean square error in the classification setting, yet so strong is the statistical intuition that it should also hold that numerous attempts have been made to generalise the bias and variance decomposition in this way [Hastie and James, 1997, James, 1998, Friedman, 1997, Kohavi and Wolpert, 1996] but there is no concensus on the best way to do so.

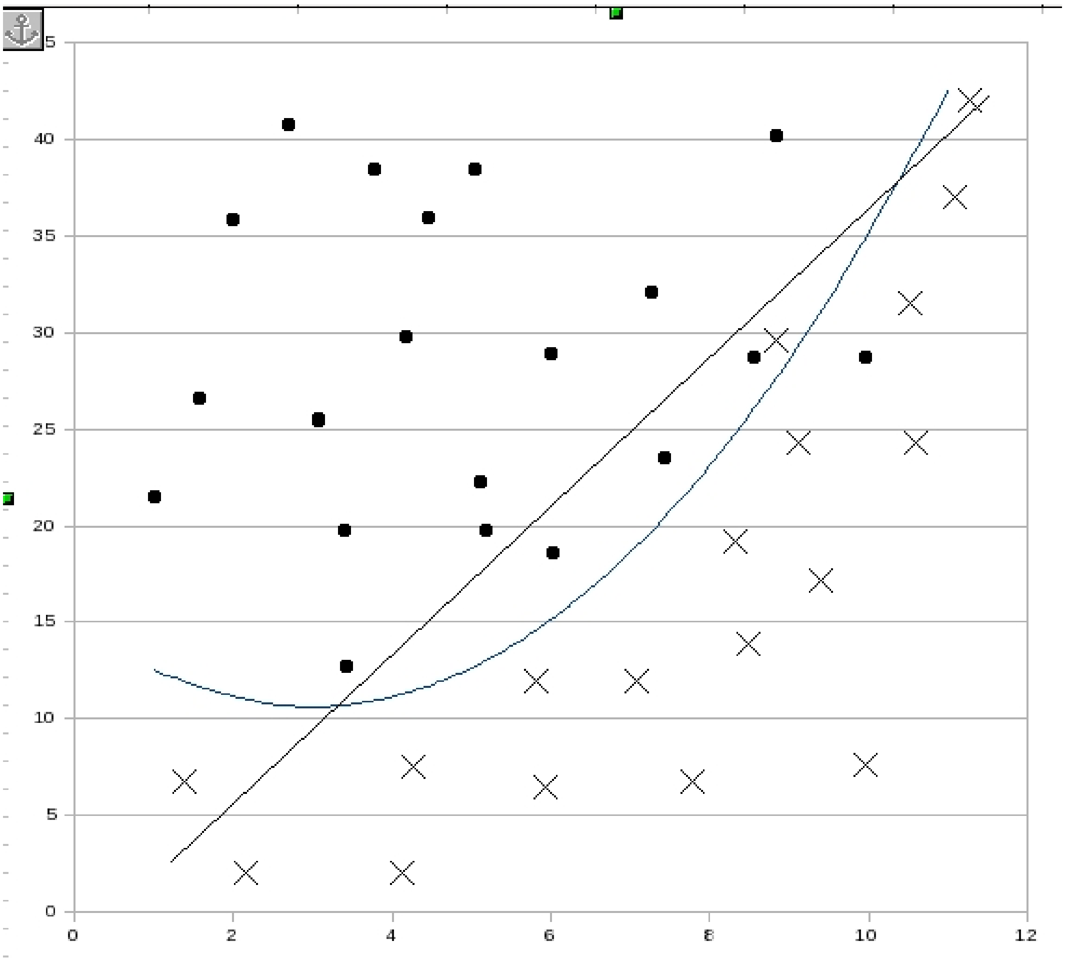


Figure 2.4: Feature space partitioned by a linear and a quadratic function.

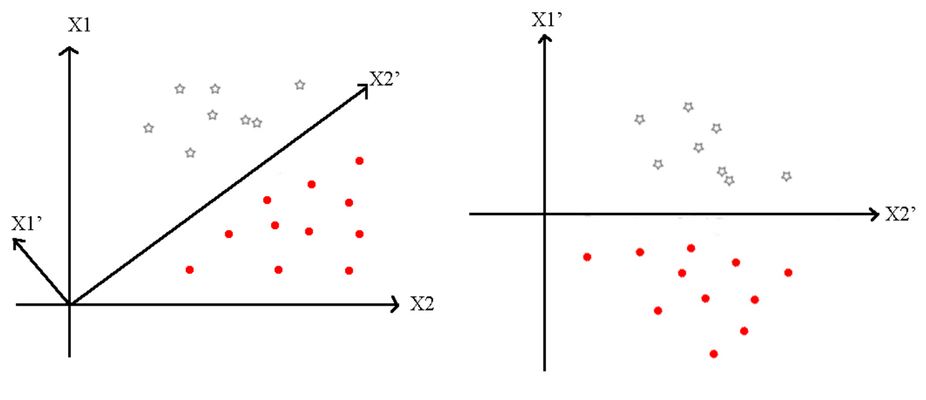
However, despite the lack of an agreed formalism, the intuition behind this search may be explored using the example of a two-dimensional decision surface. For complex data, flexible decision surfaces will reduce the inherent bias of the estimate, since they can better adapt to fit the data. However flexible rules often require more parameters to estimate than simple ones and one would expect that with too little data, the estimates for these parameters may have too much variance to be reliable. For example, consider the linear and quadratic decision surfaces in figure 2.4. For the two dimensional feature space in figure 2.4, the linear surface ($y = ax + b$) may be described by two parameters while the quadratic ($y = ax^2 + bx + c$) requires 3. As the number of dimensions increases, the difference between the number of parameters grows; at 3 dimensions, the linear surface requires 3 while the quadratic requires 6. The greater the number of parameters required the greater the individual variances of the estimates of the respective parameters, which in turn contributes to the total variance of the final estimate of the decision surface. Thus the variance of the estimate increases as the number of design samples decreases and as the number of variables increases. Furthermore, simple decision surfaces are less affected by the dimensionality and size of the design set, although they may suffer from higher bias when the optimal surface is complex.

Conventional wisdom suggests that simple decision rules may be favourable in small sample settings [Braga-Neto, 2007, p.92] and when the number of samples is very large, then more complex rules will yield better performance. This wisdom may be born out by experience and empirical evidence, but it must be interpreted with some caution. First, sometimes simple decision surfaces are the optimal decision surfaces, a good example is when the data is homoscedastic and the class distributions are normal, then there can be no better decision surface than one that is linear in the feature space (2.4.3) and more complex decision surfaces will model themselves on the noise in the data and thus impair performance. Secondly, some complex decision surfaces, like those produced by k-nearest neighbour and Learning Vector Quantisation, may be obtained using few free parameters.

Despite this, the intuition is appealing and for many real world problems, this intuition holds. That increasing the number of samples in the design set usually does lead to a reduction in the variance of the classifier has been known for a while. For example, Brain and Webb [1999] showed that on four different data sets, three popular classifiers: Multiboost, the Naive Bayes classifier, and C4.5 showed decreased variance as the sample size increased, but no increase in bias based on the definitions of bias and variance in Kohavi and Wolpert [1996]. In addition to the bias-and variance approach, there have been a number of other approaches to understanding the complex relationship between sample size, complexity of data and the complexity of the classifiers. However, even when the methods are more easily justified formally, they are less intuitive for explaining the problem.

2.3.2 Dimensionality Reduction: Principal Components Analysis

From the perspective of the feature space, two possible strategies may be adopted to handle the small sample problem. The first is to choose good features to start off with. A good set of features for classification will be one in which simple rules can easily separate the data according to respective classes. In some cases, knowledge of the problem domain can be used to construct a mapping from the space of the original measurements (sometimes called the attribute space) into another feature space in which classification may be simpler. The example (from 2.1.1) of the task of identifying the notes in an audio signal illustrates this. The task is easier when based on a frequency-power spectrum than on an amplitude-time signal.



(a) (b)

Figure 2.5: PCA illustrated. (a) is the original feature space with axes X1 and X2, axes X1' and X2' are also illustrated. (b) shows the transformed feature space which is simply achieved by rotating the axes so that the variance about X2' is at a maximum. In this case, classification may be carried out simply based on the projected values in X1'.

Even when very little is known about the data, one may also reduce the dimensionality of the feature-space by making the assumption that the variance in the data set is largely between class rather than within class.

The most popular dimensionality reduction technique is principal components analysis (PCA) in which the transformed feature space is the linear combination of the original variables in which the variance about each axis is maximised. An intuitive explanation of PCA is that it transforms the feature space by rotating and translating the original axes so that around each orthogonal axis, the variance of the data is maximised. This is easily illustrated in two dimensions. Figure 2.3.2 shows two feature spaces. X1 and X2 are the axes in the first and X1' and X2' are the axes in the second. X1' and X2' are simply X1 and X2 rotated so that the variance about the horizontal axis is maximised. Notice that only the X1' values of the data are needed for classification. Thus, the dimensionality of the feature space has been reduced from 2 to 1. This concept may be extrapolated to many dimensions. In many real-world data sets, the first few components may account for the majority of the data in the set. This has been exploited considerably in face-recognition even where the feature space is composed of the raw pixel intensities of a greyscale image (so this is 160 000 dimensions for a 400x400 image) and the number of samples per class may be as few as 20 [Turk and Pentland, 1991].

A more formal explanation of PCA follows. If $\mathcal{F}$ is the original feature-space of dimensionality n, then PCA finds the n-dimensional vector space $\bar{\mathcal{F}}$ that is a linear combination of the original variables such that for any $d < n$, the d dimensional subspace of $\mathcal{F}$ with the greatest total variance will have a basis composed of some d components of $\bar{\mathcal{F}}$. The solution to the problem is fairly simple. Let $\mathbf{x} = X_1, X_2 \ldots X_d$ be a vector of the original variables (representing the original feature space) and let $\mathbf{a_i}$ be the transformation vector for the ith new variable, then it can be shown that $\mathbf{a_i}$ must be an eigenvector of the covariance matrix (which can be approximated by the sample covariance matrix) and the variance for the transformed variable is the corresponding eigenvalue. Thus, selecting the $k < d$ variables of greatest variance means selecting the eigenvectors of the covariance matrix with the k greatest eigenvalues. There are a number of ways to derive this method, none of which are as simple as the method itself. The explanation here is due to Hotelling and follows the exposition in Webb [2002, p.322].

We seek the orthogonal variables $\xi_1 \ldots \xi_d$ that are linear combinations of $X_1 \ldots X_d$

$$\xi_i = \sum_{j=1}^{d} \mathbf{a}_{ij} X_j \tag{2.9}$$

so as to maximise the variance of ξ_1 and such that the next greatest variance of a variable orthogonal to ξ_1 is ξ_2 and so on. A maximum for any of the variances must be a stationary value, so the first step is to find these.

The additional constraint that for all $i < d$, $\mathbf{a_i}^T\mathbf{a_i} = 1$ allows the problem to be solved using Lagrange multipliers. We start by finding the stationary value of ξ_1:

$$
\begin{aligned}
\mathrm{var}(\xi_1) &= E[\xi_1^2] - E[\xi_1]^2 && (2.10)\\
&= E[\mathbf{a}_1^T\mathbf{x}\mathbf{x}^T\mathbf{a}_1] - E[\mathbf{a}_1^T]\mathbf{x}E[\mathbf{x}^T\mathbf{a}_1] && (2.11)\\
&= \mathbf{a}_1^T E[\mathbf{x}\mathbf{x}^T]\mathbf{a}_1 - \mathbf{a}_1 E[\mathbf{x}^T]E[\mathbf{x}^T]\mathbf{a}_1 && (2.12)\\
&= \mathbf{a}_1^T(E[\mathbf{x}\mathbf{x}^T] - E[\mathbf{x}]E[\mathbf{x}^T])\mathbf{a}_1 && (2.13)\\
&= \mathbf{a}_1^T\Sigma\mathbf{a}_1 && (2.14)
\end{aligned}
$$

where Σ is the covariance matrix of the vector $[X_i]; i = 1 \ldots d$. Since $\mathbf{a}^T\mathbf{a} = 1$ finding the stationary value of $\mathrm{VAR}(\xi_1)$ amounts to differentiating $\mathbf{a}^T\Sigma\mathbf{a} - \nu\mathbf{a}^T\mathbf{a}$, equating to zero and solving for ν (by the method of Lagrange multipliers) giving the eigenvalue problem $\Sigma\mathbf{a} = \nu\mathbf{a}$. Since

$$
\begin{aligned}
\xi_1 &= \mathbf{a}_1^T\Sigma\mathbf{a}_1 && (2.15)\\
&= \nu\mathbf{a}_1^T\mathbf{a}_1 && (2.16)\\
&= \nu && (2.17)\\
& && (2.18)
\end{aligned}
$$

We may select the eigenvector of the covariance matrix with the greatest value as ξ_1 guaranteeing that its variance is the greatest of all possible such variables. This variable we call the first principal component. Analogously we may define subsequent components as the eigenvectors with the next greatest eigenvalues. The derivation for subsequent components is conceptually simple but messy, and the interested reader is directed to Webb [2002, p.323].

Because PCA is not scale invariant, it is usually advisable that the original variables be of commensurate scale (unless the difference in scales is significance for classifying). This can be achieved by taking the z-scores of each variable rather than the raw scores. This also ensures that the variables are zero mean. PCA can then be conducted by finding the k eigenvectors of the covariance matrix with the largest eigenvalues and projecting the data into this subspace.

2.3.2.1 How Many Components? PCA90/70

An important issue is selecting the components to be used for classification. There are a variety of methods that are used. A popular method is to examine a plot of the eigenvalues of each component (called a skree plot) and look for a shoulder point where the variance tapers off rapidly, and then to use all the components up to the shoulder point. Another method is to consider the number of components as a free parameter of the classifier and use a validation set to select the optimal number of components empirically. Finally, because the small amount of data does not allow for an empirical method, and skree plots often lead to insurmountable difficulties with the classifiers used in this study, the method used in this study is to select the smallest number of components that account for more than a certain percentage of the variance. Throughout this study the standard is used of 90% of the variance in the case of all classifiers except for quadratic discriminant analysis in which 70% of the variance is used. The rationale is explained in the discussion of classifier types (2.4.3). This particular method of selecting components will be referred to as **PCA90/70** throughout the rest of this dissertation.

2.3.3 The Ugly Duckling Theorem

The importance of selecting good features from the start has already been stated. However, it is not a trivial task to do so based on first principles. One of the problems is that there is no universally optimal set of features - they all are heavily dependent on the nature of the problem to be solved. Thus, there is only limited value to seeking features selection and extraction strategies from the literature without carefully considering the nature of the classification task at hand. This was formally proven by Watanabi [1969, p.376] in his ugly duckling theorem which states that, given a finite set of predicates, the number of predicates that enables us to distinguish between two patterns is constant and independent of the choice of pattern.

2.3.4 Finding Features

Given the pessimistic findings of Watanabe, how can one be expected to find suitable features? The question is difficult, but it is not insurmountable. However, it is unlikely that any feature set is optimal except in very limited circumstances, such as when the process by which the data are generated in known in closed form. One approach to feature extraction is to consider which attributes are irrelevant to the classification task, and choose algorithms that are invariant to these. For example, shape descriptors are often chosen because they are invariant to rotation, scale, translation and certain transformations. However there are two caveats with this approach. First, it is impossible to exclude all possible irrelevant attributes and secondly being invariant against irrelevant attributes does not guarantee that a feature set isn't also invariant against the properties that one wishes to measure.

A second approach is to use knowledge of the problem domain. In the case of well understood systems, such as physical systems where the laws of nature are understood, this problem is usually well posed. The previous example of classifying amplitude-time signals (such as audio waves) according to their pitch, is such an example. In areas in which the problems are less well understood, useful algorithms may sometimes be found by a search through the literature for solutions to similar problems. In most cases, an appeal to scientific intuition may be required and in this study this amounts to an appeal to art historical intuition. A final approach is to use empirical evidence to test a number of feature extraction algorithms on pilot data and to use the optimal set of features as the final feature space. Each of these approaches have been used in the study, but the methods and rationale are different for each of the chapters, and thus a full discussion of the methods is deferred to the relevant chapters.

2.4 Classifier Types

The main issues surrounding the process of data collection and feature extraction have been discussed, but these do not deal with the central question of how one uses data and features to construct a decision rule. This is the subject of this section. There are many approaches to the problem and an overwhelming number of classifiers available to choose from. Throughout this dissertation, variants of linear discriminant analysis, quadratic discriminant analysis and k-nearest neighbours are used, usually in concert using a simple ensemble method (2.6). In this section these types of classifiers are described and the mathematics behind them explained and other approaches not used are described briefly. In addition, this section also provides the rationale behind the choice of classifiers used in this study.

The approach used to describe the methods is statistical, from the perspective of Bayes' theorem. Bayes' theorem is first stated and the Bayesian terminology explained (2.4.2). This framework is then used to explain the classifier types used in this dissertation. Before this, however, there is an important and often overlooked issue in classifier design that requires some discussion. That is a theorem analogous to the Ugly duckling that casts doubt on the value we place in evaluating certain classifier designs as universally better than others.

2.4.1 No Free Lunch for Learning Algorithms

A number of theorems collectively called No Free Lunch theorems (NFL) cast doubt on the possibility of ever finding a universally good classifier, where good simply means universally having better than chance performance. The NFL theorems were first stated and proved by Wolpert and MacReady regarding optimisation and search algorithms in a series of papers (Wolpert and Macready [1995, 1997]) and Wolpert [2001] expounded the implications of the theory for learning algorithms. These findings have surprising ramifications since they also suggest that even using empirical methods to selecting the best models does not guarantee better than chance performance for any learning, optimisation or search algorithm. Two theorems in particular concern pattern classification. To state the theorems we define two quantities, f, the true mapping from the feature space to the set of classes, and h, a classifier's hypothesis or estimate concerning the true f. The first NFL theorem states that the expected performance of a learning algorithm is the non-Euclidean inner product of $p(f|\mathcal{D})$ and $p(h|\mathcal{D})$ where $\mathcal{D}$ is the design set. Informally this means that the performance of a learning algorithm, including a classifier, is determined by how well it matches the true mapping and that unless there is some way of knowing what form this mapping takes, there is no way of evaluating whether one learning algorithm is better than any other. Theorem 2 is a direct consequence of this and states, informally that for all f and for all distributions over f, averaged over all possible training sets, regardless of size, the expected performance of all classifiers is the same, unless there is some prior knowledge of f.

Interpreting NFL has lead to many debates which are beyond the scope of this study. However, despite this, there are a number of consequences for the present study. First, without prior knowledge of the problem beforehand, one can neither guarantee that any single classifier will be optimal nor prove from first principles why any specific choice of classifier should perform better than chance on a specific classification task. Secondly, the notion that certain learning algorithms can be considered *a priori* superior to others is not true, although the author knows of no proven examples, artificial or otherwise, that certain algorithms (such as C4.5), and various methods of hypothesis testing for model selection (such as cross-validation), perform below chance with reasonably sized design sets.[4]

[4]Zhu and Rohwer [1996] claim a case of cross-validation performing as a bad selector between a sample mean and the median as the estimator of the true mean of a sample. However, as pointed out by Goutte [1997], while there is almost certainly no free lunch for cross-validation or any model selection algorithm, Zhu and Rohwer [1996] had not used true cross-validation and when correctly applied to the same problem, cross-validation still

In many problem domains, *a priori* information about the classification task is abundant because a lot is known about the physical systems that underlie the structure in the data. While this is also the case regarding attribution, a problem is that, unlike the hard sciences, this knowledge is not explicitly stated in such a way that it may easily be used to help model selection. Instead, this study uses art historical knowledge as much as possible, and in subtle ways, to select suitable classifier designs. For example in chapter 3, the primary assumptions are based on a well established theory of attribution - the theory of Giovanni Morelli. Chapter 4 uses knowledge about the tradition of black-figure to select the features least likely to be influenced by an artist's mood, since these are more likely to be consistent from painting to painting and thus provide a more stable basis for comparison. Of course, intuition also has to guide the choices of classifier design and these are difficult to explain and quantify. However, whatever methods have been used have been documented, and where possible scientific motivation has been provided. However, as NFL illustrates, these motivations cannot proceed from first principles.

2.4.2 Bayes' Theorem

Assume we are faced with the task of classifying an object based on a feature vector $\mathbf{x}$ to one of $C \in \mathbb{N}^+$ classes. Let ω_c be the event that the object belongs to class $c \in \{1..C\}$. A statistical approach to solving this problem would be to classify to the class c that maximises the conditional probability of ω given that feature vector $\mathbf{x}$ has been observed, or more concisely attribute to ω_j where $j = \arg\max_c p(\omega_c|\mathbf{x})$. Then the following results in elementary probability theory are of some use in seeking a solution.

$$p(\omega_c|\mathbf{x}) = \frac{p(\mathbf{x}, \omega_c)}{p(\mathbf{x})}, p(\mathbf{x}|\omega_c) = \frac{p(\mathbf{x}, \omega_c)}{p(\omega_c)} \tag{2.19}$$

$$p(\omega_c|\mathbf{x})p(\mathbf{x}) = p(\mathbf{x}|\omega_c)p(\omega_c) \tag{2.20}$$

$$p(\omega_c|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_c)p(\omega_c)}{p(\mathbf{x})} \tag{2.21}$$

Equation (2.21) is Bayes' theorem and is one of the most important formulae in pattern recognition and its frequent use in this dissertation means that it requires some explanation, and for the appropriate terminology to be developed. In equation (2.21) $p(\omega_c)$ is the probability that ω_c is the state of

performed better than chance.

nature before the observation of $\mathbf{x}$ and is referred to as the **prior** probability since it encodes information known prior to conducting the experiment or taking measurements of $\mathbf{x}$. By contrast the value of interest is the conditional probability $p(\omega_c|\mathbf{x})$, called the **posterior** probability. At it simplest, the value of the prior could represent any information that was known about the relative frequencies of the respective states of nature before the classification task was implemented. The conditional probability of the feature vector given knowledge of the class, $p(\mathbf{x}|\omega_c)$, is a fundamentally more difficult quantity to explain and to compute. It is the probability that the feature vector would be chosen at random with the particular state of nature being fixed. Finally, the denominator in (2.21) is a normalising factor that ensures that the final probability lies between 0 and 1. It can be calculated by summing over all the possible values of the numerator with respect to c. Although illustrated using simple probabilities, Bayes' theorem also holds for probability density functions and distributions.

These concepts are not intuitive and an art historical example may better illustrate these concepts. Assume a substantial fragment of a vase painting has been discovered and an expert had decided, from a close analysis of the object that it was rendered by the either painter A or painter B and the expert was 20% sure it was painter A and 80% sure it was painter B. Assume a diagnostic feature on the vase was found that appeared in 90 of the 100 paintings by A and in only 40 of the 160 paintings by B. The presence of the feature may be denoted by $\mathbf{x}$ and the possible states of nature are $\omega_1 =$ painter A and $\omega_2 =$ painter B. By Bayes' theorem the probability that painter A painted the sherd is

$$p(\omega_1|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)p(\omega_1)}{p(\mathbf{x})} \tag{2.22}$$

Here $p(\mathbf{x}|\omega_1)$ is the probability that $\mathbf{x}$ would be found on a random painting of painter A, which we estimate from history as $90/100 = 0.9$. The prior probability could be the faith of the expert in each of the attributions before he/she discovered $\mathbf{x}$ which in this case is 20% or 0.2. p(x) is simply the sum of the different values of the numerator. So for painter A this is 0.9*0.2 and for painter B this is 0.25*0.8 and $p(\mathbf{x})$ is 0.9*0.2+0.25*0.8 = 0.38. Combining these together reveals $p(\omega_1) = 0.9 * 0.2/0.38 = 0.47$ and $p(\omega_2) = 0.25 * 0.8/0.38 = 0.53$ favouring an attribution to painter B. This, in essence, is the Bayes' classifier in its simplest form. The prior probability here had a great deal of influence over the final attribution. However, one of the most useful properties of Bayes' theorem is that the posterior probabilities of 0.47 and 0.53 may be used as prior probabilities if any other diagnostic feature were found, and Bayes' theorem was again invoked for attribution.

Often a uniform prior may often be assumed and the decision rule is simply assign to the agent for which the likelihood function $p(\mathbf{x}|\omega_c)$ is the greatest. With no prior knowledge of the likelihood function, however, classification may be very difficult and two general approaches are widely adopted. Parametric approaches assume that the likelihood function follows some type or family of distribution and the task reduces to the estimation of the parameters for that distribution given a design set. Even in this case, when the distribution is complex and massively multivariate, then very sophisticated sampling techniques are required to estimate the parameters.

Non parametric approaches, on the other hand, usually aim to estimate the density using multivariate histograms taken from the training data. One of the most Draconian is the naïve Bayes classifier (NBC) in which the assumption is made that the different measurements in the feature vector are drawn from independent distributions. This is often unjustified, but the NBC has a good track record in a number of applications. Another method is to smooth the histogram by using splines or convolution with a suitable kernel. Two very important and popular parametric Bayes classifiers that are used in this study make the assumption that the parametric form for the likelihood function is normal and simply use the sample means and covariance matrices as the parameter estimates (these are the maximum likelihood point estimates for the normal parameters) These are QDA and LDA.

2.4.3 QDA and LDA

Two very important classifier types are linear and quadratic discriminant functions in which the decision surface is a hyperplane and a hyperparabola respectively. There are a number of methods of deriving linear and quadratic classifiers, but from the Bayesian perspective, these may be viewed as special cases of the Bayesian classifiers in which the likelihood function is assumed to be multivariate normally distributed. In the simplest case, where the likelihood function is normally distributed and where the data are homoscedastic (i.e. the covariance matrices for the classes are the same), the decision surface is linear and the classifier is called linear discriminant analysis (LDA). This is demonstrated in the simplified 2 class case where the assumption is made that the prior probabilities are equal and therefore assignment may be made on the basis of the ratio of the respective likelihood functions: $\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)}$, assigning the object to class 1 if the ratio is greater than 1, and to class 2 otherwise. The homoscedasticity assumption allows us to assume that the respective normal distributions have a common covariance matrix. This ratio can be expressed as the quotient of two normal distributions with means $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ and covariance matrix Σ as follows:

$$\frac{e^{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}}_\mathbf{1})^T\Sigma^{-1}(\mathbf{x}-\bar{\mathbf{x}}_1)}}{e^{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}}_\mathbf{2})^T\Sigma^{-1}(\mathbf{x}-\bar{\mathbf{x}}_2)}} \tag{2.23}$$

Since the logarithm is a monotonic function, applying it to (2.23) doesn't change which of the two functions is larger, but dramatically improves the tractability of the equation, yielding

$$\ln\left(\frac{e^{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}}_\mathbf{1})^T\Sigma^{-1}(\mathbf{x}-\bar{\mathbf{x}}_1)}}{e^{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}}_\mathbf{2})^T\Sigma^{-1}(\mathbf{x}-\bar{\mathbf{x}}_2)}}\right) = (\bar{\mathbf{x}}-\bar{\mathbf{x}}_\mathbf{1})^T\Sigma^{-1}(\bar{\mathbf{x}}-\bar{\mathbf{x}}_\mathbf{1}) - (\bar{\mathbf{x}}-\bar{\mathbf{x}}_\mathbf{2})^T\Sigma^{-1}(\bar{\mathbf{x}}-\bar{\mathbf{x}}_\mathbf{2}) \tag{2.24}$$

$$= \mathbf{x}^T\Sigma^{-1}(\bar{\mathbf{x}}_\mathbf{1}-\bar{\mathbf{x}}_2) \tag{2.25}$$

Since the discriminant function has changed from a ratio to a difference, the threshold is no longer 1 but 0 and therefore assignment can be based on whether the discrimant function is positive or negative.

Removing the homoscedasticity requirement, if we represent the covariance matrix of class 1 by Σ_1 and the covariance matrix of class 2 by Σ_2 then for the two class case the Bayes classifier yields the following discriminant function (following Duda et al. [2000]):

$$\ln(\Sigma_1^{-1}) + (\mathbf{x}-\bar{\mathbf{x}_1})^T\Sigma_1^{-1}(\mathbf{x}-\bar{\mathbf{x}}_\mathbf{1}) - \ln(\Sigma_2^{-2}) + (\mathbf{x}-\bar{\mathbf{x}_2})^T\Sigma_2^{-2}(\mathbf{x}-\bar{\mathbf{x}}_\mathbf{2}) \tag{2.26}$$

which is a quadratic function and the classifier is sometimes called quadratic discriminant analysis or **QDA**. Typically the flexible QDA tends to overfit the classification rule to the design set in small sample cases, particularly if the data is not actually normally distributed. In fact, even if the data are normally distributed and heteroscedastic, LDA and nearest neighbour methods discussed below often yield better results when data are scant.

Both LDA and QDA may be generalised to the multiclass case and to the case in which prior probabilities are specified. This time, we define a discriminant function g_i for each class i which is simply the log of the posterior distribution (again, the logarithm is used because it makes the equations more tractable):

$$g_i(\mathbf{x}) = \ln\frac{p(\mathbf{x}|\omega_i)p(\omega_i)}{\text{normalising factor}} \tag{2.27}$$

Here $p(\omega_i)$ is the prior probability that agent i is responsible for the form, and the normalising factor is independent of the class and may be ignored.

Thus, because the likelihood is a normal distribution, $g_i(\mathbf{x}) = \ln(\text{prior}) + \ln(\text{normal distribution})$ which in turn gives:

$$g_i(\mathbf{x}) = \ln(p(\omega_i)) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_i)^T\Sigma^{-1}(\mathbf{x} - \bar{\mathbf{x}}_i) \quad (2.28)$$

The term $-\frac{d}{2}\ln(2\pi)$ may be ignored because it is the same for each class. In LDA, because of the homoscedasticity assumption, the term $\ln|\Sigma_i|$ may also be omitted and furthermore, the term $\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_i)^T\Sigma^{-1}(\mathbf{x} - \bar{\mathbf{x}}_i)$ may be further factorised (a messy process which is omitted here) finally resulting in $g_i = (\Sigma^{-1}\bar{\mathbf{x}}_i)^T\mathbf{x} - \frac{1}{2}\bar{\mathbf{x}}_i^T\Sigma^{-1}\bar{\mathbf{x}}_i + \ln(p(\omega_i))$. In QDA, terms involving Σ_i cannot be ignored and the resulting discriminant function is more hairy:

$$g_i = \frac{1}{2}\mathbf{x}^T\Sigma_i^{-1}\mathbf{x} + (\Sigma_i^{-1}\bar{\mathbf{x}}_i)^T\mathbf{x} - \frac{1}{2}\bar{\mathbf{x}}_{\mathbf{i}}^T\Sigma_i^{-1}\bar{\mathbf{x}}_{\mathbf{i}} - \frac{1}{2}\ln|\Sigma_i| + \ln(p(\omega_i)) \quad (2.29)$$

Assignment is then made to the class for which g is the highest. Both LDA and QDA may be implemented such that the posterior probability for membership to each class is also reported - which is simply the exponent of g_i (although without the superfluous terms removed). In this study, both LDA and QDA are implemented using MATLAB's classify() function.

2.4.3.1 Multiple Discriminants Analysis and Canonical Variables

A second approach to LDA is the classical approach due to Fisher, commonly called Fisher's Linear Discriminant Analysis (FLDA). The aim of FLDA is to find a component that is a linear combination of the original variables that, in some sense, maximises the difference between the classes. Fisher's approach is to define a measure of dispersion between the elements of each class with elements of the other classes and compare this with a measure of dispersion between elements within the same classes. On this basis, Fisher's LDA seeks to find the direction of the vector that most maximises the ratio of the between class dispersion and the within class dispersion.

The reason Fisher's method is discussed here, despite the fact that this study uses the Bayes normal linear classifier, is that the multi-class extension of Fisher's method, multiple discriminant analysis (**MDA**) may be used to seek a lower dimensional subspace that most accounts for the separation between classes. These variables are referred to in this study as the **canonical variables** and projecting the data into the subspace spanned by S canonical variables with the highest associated eigenvalues is equivalent to representing the data in the S dimensional subspace that is in some sense optimal for discrimination between the classes. In 4.2.4.2 the canonical variables are used to investigate the structure of the data.

The exposition of MDA follows Tebbens and Schlesinger [2007]. First, if we denote the design set by $\mathcal{D}$ then C subsets of the design set may be defined, where C is the number of classes: $\mathcal{D}_i, i = 1..C \in \mathbb{N}^+$ such that $\mathcal{D}_i$ contains all n_i members of $\mathcal{D}$ belonging to class i. If N_i denotes the index set of members of class i then the within class and between class covariance matrices, $\mathbf{W}$ and $\mathbf{B}$ respectively are defined as

$$B = \frac{1}{C-1}\sum_{i=1}^{C} n_i(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \tag{2.30}$$

$$W = \frac{1}{n-C}\sum_{i=1}^{C}\sum_{j\in N_i} (\mathbf{x}_j - \bar{\mathbf{x}}_i)(\mathbf{x}_j - \bar{\mathbf{x}}_i)^T \tag{2.31}$$

MDA seeks a set of $S < C$ vectors $w_j, j = 1 \dots S$ that maximise the quotient $J(\mathbf{w}_j) = \frac{\mathbf{w}_j^T B \mathbf{w}_j}{\mathbf{w}_j^T W \mathbf{w}_j}$. The solutions may be found by solving the generalised eigenvalue problem $(B - \lambda W)\mathbf{w}_j = 0$. The S eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$ with the highest eigenvalues will span the S dimensional space that most separates the different classes in the sense discussed above. If S = 1, the $\mathbf{w}$ with the highest eigenvalue is the direction of the optimal component for classification if the number of classes is 2.

QDA and both methods of deriving a linear discriminant require inversion of covariance matrices, and this is only possible if the rank of the covariance matrices is greater than the number of observations. In many small sample problems, this condition is not met and methods have to be devised to circumvent the problem. Numerous examples have been given. A recent summary [Tebbens and Schlesinger, 2007] lists perturbation methods that add small values to the singular values, the use of the Moore-Penrose pseudo-inverse and methods that concentrate on the nullspaces of the respective covariance matrices. In this dissertation dimensionality reduction, in particular PCA, is used to reduce the dimensionality of the feature space and thus avoid singular covariance matrices. This method has come under criticism for not preserving important discrimination information that may be contained in the nullspace, but the literature shows an impressive track record for this method when applied to small sample studies

The method used in this study to determine how many components to use (PCA70/90) has already been described. Briefly, the smallest number of components that account for 90% of the variance is used except for QDA in which the cutoff is 70%. The reason for the difference is that LDA inverts the pooled covariance matrix, while QDA inverts the covariance matrix for each class. Because the lower components contain most of the variance in the data, selecting too many components increases the risk of the covariance matrix being singular. Since the pooled covariance matrix has more data than the individual class covariance matrices, more components may typically be used before it becomes singular.

2.4.4 Direct Posterior Estimation

Some algorithms directly estimate the posterior class conditional densities. Examples of this kind of algorithm include Parzen windows and k-nearest neighbour (**k-nn**) methods. The former selects a window with a specified hyper-volume and calculates the conditional densities for all areas of the feature-space. The chief difficulty here is determining the optimal size of the window, which is usually done during the validation phase. Without good selection of window size, Parzen windows are prone to poor performance in high dimensional feature spaces.

A simpler method to overcome the issue of the size of the windows is to work in reverse by specifying a number of samples and, centring on each sample, estimating the minimum hypervolume required to cover this number of samples. This allows a class conditional density to be estimated for the areas around each sample (rather than for the whole feature-space). This is the idea behind the k-nearest neighbour method (k-nn) which simply classifies each feature vector to the same class as the majority of the k feature vectors in the training set that are nearest to it according to some metric defined on the feature space (often Euclidian distance can be used). Although there are certain difficulties in using k-nn to estimate densities, it is particularly simple in classification tasks, since one merely assigns any given object to the class to which most of its k neighbours belong. Adjusting the parameter k effectively alters the complexity of the decision surface - higher values of k produce smoother decision surfaces, lower variances and higher biases, while low values of k produce very detailed decision surfaces, high variance and low bias. In fact, when k is 1 and the sample number is large, k-nn is virtually unbiased, but the variance of the classifier is large. Thus, by altering k, one easily adapt the algorithm to problems of various sample sizes - including small sample studies. Finding the optimal value of k may be done during the validation phase (2.5). For k-nn, the feature space or the metric used to measure the distance between neighbours is vital to the method's success. There are many methods for finding a suitable metric, but one that works well in small sample setting and the method that is used in this study to use PCA to find a small number of variables and apply k-nn to this feature space using Euclidean distance as the metric. In this study, k-nn is implemented using MATLAB's knnclassify() function.

2.4.4.1 Methods Used in this Study

The classifiers mentioned here are only a few of the approaches that have been developed. All have their strengths and weaknesses and by NFL, there can be no purely principled reason to chose one over any other. LDA is most commonly used in this dissertation for a number of reasons. First, the linear discriminant function is easily tractable allowing a greater understanding of the problem space than with a less tractable models like neural networks. Secondly, and most importantly, the linear classifier, with its rigid assumptions, has very high bias, but consequently low variance. As has been argued, this can be beneficial in small sample studies (2.3.1.1). LDA, despite its simplicity has achieved very impressive results in most real world benchmarks, even when applied to tasks to which it is not suited (for example, Lim et al. [2000] find LDA and a variant of LDA: LOG (Logistic Discriminants analysis) both consistently performed in the top 5 out of 34 classifiers on a variety of tests). In addition to LDA, nearest neighbours and QDA are used in ensembles but a motivation for this is deferred till ensemble methods have been explained in more detail in 2.6.

2.5 Performance Evaluation

There are two stages in the design cycle of a classifier where performance evaluation are likely to be used: model selection and evaluation. Model selection is the process of choosing between a number of candidate classifier models for a classification task. Evaluation is the process by which the performance of the classifier on novel instances (i.e. real world examples it has not encountered in training or validation/model selection) is estimated. In this study, the term **validation** will be used to describe a particular kind of model selection in which the free parameters of a classifier design are selected so as to optimise the performance of the classifier. For example the value of k in k-nn or the number of components in PCA + LDA may be chosen so as to optimise the performance of the classifier on new data. This may be achieved in many ways, all of which either implicitly or explicitly estimate the relative performance of the classifier under the respective combinations of parameters.

A popular, non optimal, approach to solving this problem is to treat the parameter selection as a constrained optimisation problem over a bounded continuous function of the parameters which is sampled at certain points (using the validation set). Gradient descent or some other numerical method may then be used to approximate a global maximum. This approach does not work when the function of the parameters is clearly piece wise continuous, in which case bounded optimisation procedures will fail between discontinuities. Such is the case when attempting to determine the best k in k nearest neighbour (where k is an integer). Furthermore, in small sample cases, it is often not feasible to use gradient descent. There are a wide variety of methods used to evaluate different models, include the Bayes Information Criterion (BIC), Minimum Description Length (MDL) and the Aikike Information Criterion (AIC), all of which evaluate the performance and penalise overly complex models - in other words they choose the model that provides a combination of the best performance and the least complexity. The interested reader is directed to Buckland et al. [1997] who provides a discussion and evaluation of these different approaches.

A popular alternative is to use empirical methods of some sort to estimate the performance of the classifier under different choices of free parameters. These empirical methods are the same as the methods that are traditionally used for performance evaluation so they are discussed as one. In this dissertation, validation is used only in chapter 3 as part of one of the approaches, and chapter 5 uses a novel method for model selection that is explained in that section. Because of the small samples, a technique often used for validation called cross-validation, has been used for performance evaluation in this study. The exact details follow in 2.5.2.

2.5.1 The Holdout Method and Mean Error Rate

The holdout method is the approach most often adopted when large amounts of data are available. In this case, the data may be partitioned into separate test, validation and design sets. The performance is measured on the test set exclusively. When there is not sufficient data for separate sets, then different strategies must be employed. If there is sufficient data for a design and test set, but not for a validation set, the design set may also be used for validation by re-using the data. While using the design set for both validation and training may result in overfitting, there are a number of methods for mitigating this effect, including cross validation and resubstitution.

Together with the choice of method of partitioning the data set, the criteria for performance measurement is the most important issue in both validation and error estimation. A widely used measure of performance is mean error rate. In particular, since the performance on novel instances is of interest, we make a distinction between test error and true error (the latter will be referred to by ε throughout this study). Test error is the empirical measure of error on the test set. For example if the test set has 30 samples of which 20 are correctly classified, then the test error is 1/3. True error is rate at which the classifier will mis-classify novel instances.

Before discussing methods of getting estimates of the true error, it is important to point out that simply reporting the test error is insufficient for most purposes since it does not tell us how much faith we should have in the estimate. It should be intuitively obvious that the more samples with which to infer a statistic, the more faith we may have that the statistic is accurate. For example if a classifier correctly classifies 2 out of 2 instances, no-one would be brave enough to suggest that the classifier's true error is 0. However, even though it is unlikely, if a classifier correctly classifies 100 000 of 100 000 instances we would be more willing to concede that in practice the classifier's true error was as good as 0. In traditional hypothesis testing, this is usually done by means of confidence intervals which may also be used when inferring the error of a classifier.

There are at least two different approaches to obtaining the true error from the test error and the method chosen depends partially on the philosophical leaning of the experimenter and partially on the nature of the problem. The traditional method is to use the test error as an estimate of the true error. Assume that the test set is composed of N samples out of which the classifier misclassifies k. Then the test error is simply the number of samples missclassified, k divided by N. Since without knowing which of the samples were misclassified, the probability of misclassification is the same for all the samples in the test set, and consequently, k is binomially distributed. If the number of samples is large, the test error is a good approximation to the true error, i.e $\hat{\varepsilon} = k/N$ and we can trust the confidence intervals established on this by assuming that k/N is normally distributed (which is a reasonable assumption when N is large).

With small samples, traditional confidence intervals become inappropriate for two reasons. First, the normal distribution can no longer serve as a good basis to establish confidence intervals and the binomial distribution must be used instead. Secondly, the variance of the binomial distribution is $N\varepsilon(1-\varepsilon)$ and as the test error approaches 0 (where ε is the probability of misclassification, or the true error), the variance approaches zero. Consequently, if a classifier makes no errors on a test set (which is quite plausible if the test set is small) the confidence interval will be of zero length, implying that the true error is *precisely* 0. It is intuitively obvious that this is estimate must be erroneous.

Now the statistic that we are interested in is the true error or ε. In particular, the Bayesian method assumes that the true error is a random variable, and like all random variables it has a distribution. Therefore if we can find the distribution we have a single function that describes not only a single value for the error, but also a measure of how much faith we should have in that single value. For example, figure 2.6 shows distributions describing the true error of a hypothetical classification rule. The estimates for the error are the modes, where the distribution peaks (in this example 0.2). The width of the distribution gives us a measure of the precision of that estimate. A narrower distribution means higher precision and the figure illustrates that the greater the number of samples in the test set, the more precise the estimate.

Reporting an entire distribution is cumbersome, so some simple statistics may be used to describe the width of the distribution. One is the variance, but exactly what the variance tells you about the width of a distribution depends to some extent on the type of distribution. A second is credible intervals which are like confidence intervals but based directly on the random variable. There are many ways of describing credible intervals, but a common one is to use a lower and an upper bound for the estimate for which we are 95% certain contains the true value of the parameter of interest. In our case, since only proof of concept is important, it is reasonable to simply report an upper bound which we are 95% certain is above the true error. Very precise estimates of the classifiers' performance in the real world, while desirable, is not of paramount importance in this case since it is sufficient to show that the results are unlikely to be due to chance. If the error expected by chance is 0.5 then intuition would suggest that all three of the distributions in figure 2.6 represent classifiers that perform better than chance since the vast majority of the areas under the curves lie below 0.5, even where only 10 samples were used. However, even in the case where the number of samples is 1000, the distribution is not narrow enough for us to have much faith that the true error is exactly 0.2.

To derive a method for calculating the distribution of the true error, we state formally what we are seeking. This is the probability distribution of the error given the data. The data in this case is the value of k and N in a Bernoulli trial (i.e. a sequence of experiments each of which have one of two outcomes that occur with a single probability - such as classify/misclassify). We summarise this as $p(\varepsilon|Bi_{k,N})$. Using Bayes theorem we may calculate this posterior probability as being equal to $p(Bi_{k,N}|\varepsilon)p(\varepsilon)$/normalising factor where $p(\varepsilon)$ is a distribution that expresses our prior belief in the possible values of the error before we saw the data.

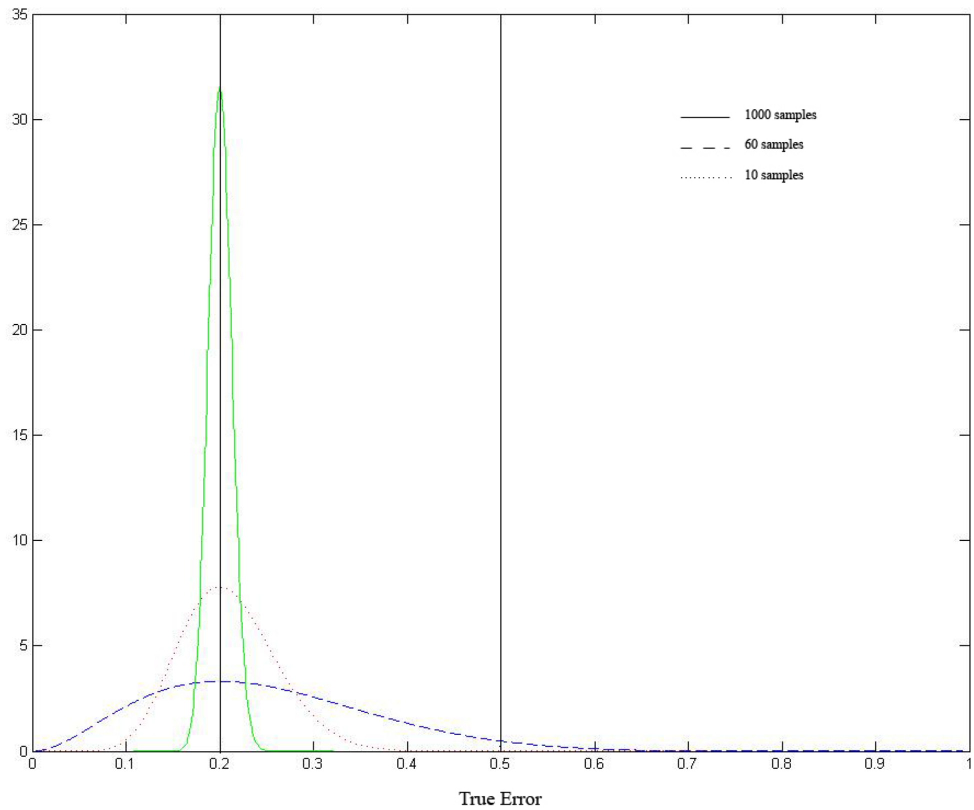


Figure 2.6: Three examples of beta-distributed estimates of the true error. The value 0.2 is the actual point estimate of the true error (the mode). However, the precision of this estimate increases as the number of samples increases. The vertical line in the middle represents the error expected by chance for a 2 way classification.

$$p(\varepsilon|Bi_{N,x}) = \frac{p(Bi_{n,k}|\varepsilon)p(\varepsilon)}{\int_0^1 p(Bi_{n,k}|\varepsilon)p(\varepsilon)d\varepsilon} \tag{2.32}$$

$$= \frac{\frac{N!}{k!(N-K)!}\varepsilon(1-\varepsilon)^{N-k}p(\varepsilon)}{\int_0^1 \frac{N!}{k!(N-k)!}\varepsilon(1-\varepsilon)^{N-k}p(\varepsilon)d\varepsilon} \tag{2.33}$$

$$= \frac{\varepsilon(1-\varepsilon)^{N-K}}{\int_0^1 \varepsilon(1-\varepsilon)^{N-k}d\varepsilon} \tag{2.34}$$

Setting $\alpha = k + 1$ and $\beta = N - k + 1$ (2.34) is beta-distributed with parameters α and β). From the posterior distribution for the error, the true error may be estimated either by the mean of the distribution which is $\alpha + 1/(\alpha + \beta + 2)$ or the mode, which is the peak of the beta distribution. It is a simple matter from the cumulative distribution function of the beta distribution to calculate the 95% upper bound which has a very simple and intuitive interpretation (unlike the confidence interval which is often misinterpreted, even by scientists). The 95% upper bound of the error estimate means that given absolutely no prior knowledge of the error rate, we may be 95% certain that the true error is bounded above by this value.

2.5.2 Resubstitution in Small Sample Settings

When the available data set is small there may not be enough data to hold-out data for a separate validation or even test set. In these cases, re-using the training data for performance evaluation may be required. The naïve approach is to use the **resubstition estimate** (resub), in which the classifier is evaluated on its performance on the training set itself. Obviously resub is a very optimistic estimate of the error which, if used for evaluation of the classifier would be misleading and if used for validation will result in overfitting and poor performance. The classical approach to solving this problem is v-fold **cross validation** (c-val) in which the training set is split into v equal sized partitions. The classifier is trained on v-1 of these partitions and tested on the remaining partition, and this process is repeated until all v partitions have been tested and the final error estimate is the combined error on all v partitions. The extreme case of this is the **leave-one out** (LOO) estimator in which only one sample is left out of the training set and the process is repeated n times where n is the number of training samples. Very closely related is the **jackknife** estimate in which cross-validation is used to estimate the bias of the resubstition error and this in turn is used to estimate the true error. The methods described here have been criticised for high variance, and furthermore, c-val and LOO are slightly pessimistically biased because the entire training set is never used.

The **bootstrap** estimator has been used to approximate the distribution of any statistic for which no closed form solution to the distribution exists (such as the median or mode of a distribution), and the same method may be used to estimate error rate. The bootstrap works by resampling the design set with replacement and creating a number of bootstrap sets of the same size as the design set. Each of these bootstrap sets is likely to have has some data repeated in the set, and is conversely unlikely to have all the data of the design set. For each set, the classifier is trained on the data in the set, and tested on the data that is not in the set. The mean value of the error rates is the estimate of the true error, and the variance between these different error estimates is the variance of the final bootstrap estimate.

A method of improving on the bias of the resub error estimate and on the variance of c-val is bolstered error estimation [Brago-Neto and Dougherty, 2004] which has outperformed many of the methods so far discussed in a wide variety of simulations of small sample problems. Of course, this is no guarantee that it will perform better on all data sets. However, in the case of LDA, bolstered error estimation can be shown analytically to be a better estimate than LOO or c-val, and for this reason has been used in this study. Bolstered error estimation takes a variety of forms, all of which have in common that instead of treating each feature vector as a single point, they treat the vector as the peak of a multivariate normal distribution. Then error may be calculated not only on the basis of how many points lie on either side of the decision surface, but also on the proportion of the normal distributions surrounding each point that is on either side of the decision surfact. In chapter 3, bolstered leave-one out is used for validation, as a method of selecting the optimal number of components (3.4.2). The idea is exactly the same as LOO, but where the feature vectors are treated as multivariate normals rather than points in the feature space. The idea is illustrated in figure 2.5.2. The reason BLOO is used in this cased is because when selecting the optimal number of components for classification using LOO, one often gets tied scores when the sample size is small, but this is not the case when BLOO is used.

2.5.3 Alternatives to Error Rate

Although accuracy and error rate are still very popular, their applicability as a measure of future performance is often criticised for not reflecting the classifier's performance in the real world [Provost et al., 1998]. First, it assumes that the costs of misclassification are equal among all classes when

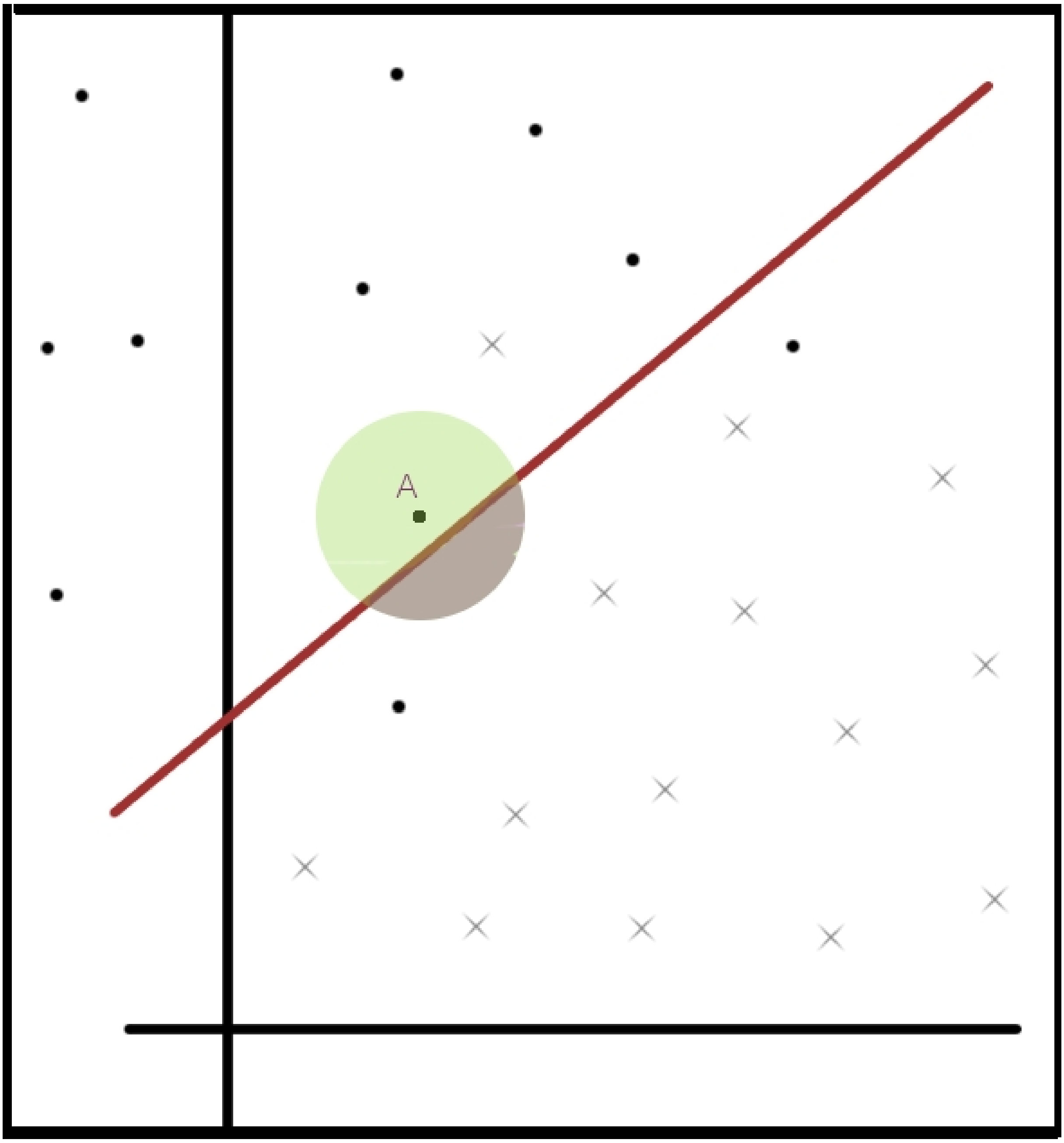


Figure 2.7: Illustration of bolstered error estimation. The points and crosses are considered to be the peaks of 2D distributions illustrated by the circle around point A. The error estimate is not based on the number of items on either side of the decision surface, but on the proportion of the normal distributions on either side.

this is often not the case. For example, in medicine, diagnosing a healthy patient with a serious disease is usually less serious than failing to diagnosing a sick patient. The second problem is that it is only accurate if the proportion of objects in each test set matches the proportion of these objects in the general population. For example if a classifier can always recognise an object of type A, but 50% of the time misclassifies an object of type B as an object of type A, then if the test set is composed of an equal number of type A and B objects, the accuracy will be estimated as 75% (assuming a large test set). However, if type B objects predominate in the real world, then 75% will be an optimistically biased estimate.

One approach to solving this problem is also to report confusion matrices. If C is the number of classes then a confusion matrix $\mathbf{M}$ is a CxC matrix in which the rows represent true classes, the columns predicted classes and the element $\mathbf{M}_{ij}$ represents the number of objects that are actually class i but are classified into class j by the classifier. Confusion matrices are useful for addressing the problems mentioned above but are also informative in cases where misclassification may reveal structure in the data.[5] Confusion matrices also form the basis of ROC analysis [Fawcett, 2006], which is a technique that has gained popularity over the last decade but which is most useful when there is both an abundance of data. In addition, it is not immediately obvious how to use ROC analysis in the multi-class case, since it is designed for two class cases.

2.5.3.1 Methods employed in this study

However, error is a perfectly valid measure for the purposes of our study for a number of reasons. First, real world proportions of the different painters in all the experiments in this study cannot be ascertained since we cannot be certain that the proportion of extant objects surviving is not due to chance rather than being reflective of the likelihood that any new object may belong to a particular class. Secondly, in most of the studies, the accuracy is quite uniform throughout the classes. When this is not the case, confusion matrices are used, particularly since these may sometimes indicate whether the classifier is indeed mimicking human intuitions about the closeness between classes.

Finally, it should be reiterated that the purpose of this study is proof of concept rather than to show the superiority of the classification rules or classifier designs over others (since these are the first supervised learning algorithms to be applied in this specific problem domain). This does mean that

[5] Such as in 4.2.4.1 where it appears that painters who are stylistically similar by art historical reckoning are confused more often by the classifier.

the statistical statement being made in each study is not that our estimate of the classifier performance is an accurate assessment of the classifier's real world performance, but just that we may reject the notion that the performance is due than chance. Given the sample constraints, the method used in this study to estimate error is by the Bayesian method already discussed in which the mode of the posterior distribution of the error is reported together with the 95% upper bound. In all three subsequent chapters, the error is estimated using some variant of the leave one out method.

For most of the study, the need for validation is entirely obviated by the use of ensemble methods in which all models are considered and used in concert to produce a final classification rule. Using all models considered has a number of benefits. First, there is no need to hold out important data for validation in a study in which data is very scarce. Secondly using all the models considered is a transparent method that does not allow the user to try multiple models and simply report the optimal results. Finally, as is explained in the next section, the ensemble methods, under the right circumstances, can outperform each of the individual models considered.

2.6 Combining Classifications: Ensemble Methods

Often when given an important diagnosis, patients seek a second opinion. The intuition that underlies this behaviour seems to be the belief that two expert opinions are somehow better than one, and that it is unlikely that two doctors who have not conferred about the case, will come to the same unlikely conclusion by chance alone. The same intuitively may be applied to classifier design. It seems obvious that if different classifiers are treated as experts, then having the "opinion" of more than one of them should somehow be better. Two questions logically follow. First, is there any principled reason to believe this to be the case, and if so, what are the best rules with which to combine the output of multiple classifiers? In answer to the first, Condorcet's jury theorem states that if each of the jurors in a jury is more likely than not to come up with the correct verdict, and each juror come up with their judgement independently, then as the number of jurors increases, so does the probability that the correct verdict will be reached [Polikar, 2006, p.34]. This result suggests a system in which multiple classifiers each cast a vote as to the class membership of an item, and then assignment to a particular class is according to a majority vote, may well perform better than the individual classifiers themselves. This is the simplest example of a majority vote ensemble, and the individual classifiers are referred to as **base classifiers**. For example, if 3 diagnostic techniques for an illness are available, then one way of achieving a diagnosis is simply according to whichever diagnosis at least 2 of the 3 techniques predict.

The conditions under which we would expect the majority vote to be better than the individual predictions of the base classifiers are analogous to those of the jury theorem: that each classifier is more likely than not to make a correct prediction, and that the errors made by each classifier are independent. To try and ensure this, many methods have been used to try and maximise diversity between the classifiers in an ensemble. There are many measures of diversity and the subject is an active area of research for which a critical summary may be found in Aksela and Laaksonen [2006].

2.6.1 Perturbing the Design Set

Dietterich [2002]'s brief overview of ensemble methods cites a number of methods that may be used to try and get the classifiers to be diverse. Perhaps the most popular is by perturbation of the design sets in repeated trials, and then taking a majority vote of the predictions made in each trial. **Bagging** [Breiman, 1996], an abbreviation of bootstrap aggregation is one of the simplest. The idea is that a number of different bootstrap design sets are sampled with replacement from the training data. That is, each bootstrap set is the same size as the design set but because the examples are sampled with replacement, some examples appear more than once in a bootstrap set, and other do not appear at all. If many bootstrap sets are created, it is likely that each member of the design set will appear at least once in at least one of the bootstrap samples. A component classifier is trained on each of the bootstrap design sets and the predictions on the test set are combined using a majority vote (i.e. assign each example to the class which is predicted most often among all the component classifiers).

Of course, bagging cannot guarantee that the component classifiers are independent, one of the conditions under which the jury theorem holds. In order to maximise independence, the component classifiers should ideally be likely to make different mistakes so that the mistakes are averaged out. One approach to achieving this is to modify the bagging algorithm by using different classification rules for each of the component classifiers. Other ensemble methods take this into account by indirectly weighting the different classifiers. A very popular method, **arcing**, achieves this by repeatedly constructing a bootstrap design set. Each time a bootstrap design set is constructed, it is tested on the design set (for example using cross-validation) and the empirical distribution is modified such that the samples from the design set that were correctly classified are less likely to appear in subsequent bootstrap samples. The process is repeated many times, and the predictions of the component classifiers are again combined using a majority vote.

Similar to this is **boosting** [Schapire, 2003], of which a number of variants exist, and which is regarded as one of the most important discoveries in machine learning in recent years. The original boosting algorithm creates three component classifiers sequentially: C1, C2 and C3. The first, C1, is trained on a random subset of the data. The second, a subset of the data used in C1 for which C1 misclassified half of the samples is used to train C2. Finally C3 is trained on instances in which C2 and C1 disagree. Classification of a novel sample is to the class to which at least 2 of the 3 component classifiers have assigned it.

2.6.2 Training on Different Features

Another approach to constructing diverse base classifiers is to train different sets of classifiers on different areas of the feature space. In other words, one may train different classifiers on different measurements of the same data. There are many different approaches to this, including many that select different random subsets of the feature space and use an intelligent algorithm to select the components based on accuracy and diversity. However, such methods work best if there is enough data for model selection, which is not generally the case in this dissertation.

Instead, for the most part in this dissertation, the base classifiers are trained on features obtained by different combinations of primary feature and secondary feature extractors. The reasoning is that if the different feature extractors are based on different principles and there is consequently little prior reason to believe that they are highly correlated, then we should expect the classifiers trained on them to be diverse. The method is similar to that used by Cherkauer [1996] to label volcanoes on Venus. Chaukauer used features derived from PCA, FFT and other secondary extractors taken from images that had been labelled by an expert and used this to train a number of neural networks. The closest analogy to Chaukauer's method in this dissertation is Chapter 5 in which 75 different feature spaces are constructed based on a variety of primary and secondary feature extractors in combination. In chapter 3, 5 different feature extractors are used and 3 different classifiers are trained on each of the 5 resulting in 15 classification rules. This is compared with the individual base classifiers to establish their effectiveness. In chapter 4, only one primary feature extractor is used, but from it, 4 different feature spaces are constructed and 3 classifiers each trained on the 4 spaces, resulting in 12 different classification rules. In each of the chapters these are combined using a majority vote without any weighting. In other words, for each item to be classified, each base classification rule makes a prediction for class membership and assignment is made to whichever class is voted for most often, with no weighting.

2.6.3 Why do Ensembles Work?

It has became obvious from empirical studies that ensemble methods do indeed work and some effort has been made on the part of the machine learning and statistics communities to understand why. Breiman [1998] showed that with bagging and arcing, multiple classifiers have little effect on the bias of the individual classifiers in the class, but the variance of the ensemble is considerably less than that of the individual classifiers. He proposes that unstable classifiers be used in the ensemble since they are most affected by perturbation of the design set. In particular, he showed that LDA, a very stable classifier, was not improved by bagging. However, because the design set is not perturbed in the present study, LDA is used because it has low variance and often makes very accurate predictions with small samples. However, to increase diversity, in chapter 3 and 4, it is used in combination with less stable classifiers QDA and nearest neighbour. In chapter 5 only LDA is used because the number of independently chosen feature spaces is so large that diversity is almost ensured.

As has been suggested earlier (2.3.1.1) in small sample cases, variance should be attacked quite aggressively. In this case, we assume that increasing the number of variables increases the discriminating power of the classifier. However, there is a consequent increase in the variance of the classifier and this negatively impacts on the performance which in the small sample case may be significant. Thus by reducing the variance using ensemble methods, we are able to use the added information in the more complex feature space without incurring the penalty of increased variance. Breiman's interpretation is not uncontroversial since boosting has been shown to increase classification accuracy even in cases where it increases variance [Schapire et al., 1988]. A more complex analysis has been suggested which explains the performance of majority vote ensembles in terms of margins, which may intuitively be described as the confidence of a classification. The suggestion is that ensembles increase the margin and thus even when variance is increased, the classification accuracy is not impaired. The argument is complex and unsolved since Breiman [1998] has illustrated cases in which margins may be decreased by ensembles even when there is an attendant increase in accuracy. We assume, that in most of our cases, using the majority vote decreases variance in addition to whatever benefits may be achieved by an increase in the margin and for this reason their use is justified in the small-sample environment of this study. Moreover, as has been explained in the previous section, the way ensemble methods are used in this study obviates the need for validation and model selection.

2.7 Conclusion

2.7.1 Summary

This chapter presented the chief theoretical concepts, methods and terminology employed in this study. This study presents attribution as a pattern recognition topic. We define the concept of a **form** which is a physical part of a painting or group of paintings that can be expressed numerically. The form is represented by a **feature vector**. Functions that convert forms into feature vectors are termed **primary feature extractors**. The space of all possible values of a feature vector is called a **feature space**. There are occasions when a property of a form may be revealed in one feature-space but not another. Thus it may sometimes be useful to transform one feature space into another. A function that does this is called a **secondary feature extractor**.

Every form must have been produced by some stylistic entity. Typically this would be a painter, but both because we are unsure that the artistic personalities discovered by Beazley were necessarily individual painters, and because the attribution task is sometimes to identify the art movement, the period or a select group of artists, this study uses the more neutral term **agent** to refer to any of these artistic personalities, be they painters, groups of painters, art movements, etc.

The attribution process may thus be expressed as follows. Some piece of artistic expression is presented for attribution. This is a form. A set of candidate agents is selected as plausible attributions. A set of exemplars of the respective agents' forms must be collected with due care taken that apples are compared with apples. These exemplars, together with the form to be attributed are measured digitally by a primary feature extractor and possible some secondary feature extractors to produce a set of feature vectors that describe the form in sufficient detail for a classification to be attempted. This set of feature vectors for the exemplars is called the **design set** and this is used to train a **classifier** to produce and implement a **classification rule** which predicts for a novel feature vector the candidate agent most likely to have produced the form described by the particular feature vector. When the feature vector associated with the form to be attributed is presented to the classification rule, the output is the attribution.

This study does not actually conduct attributions of novel examples. Instead it is concerned with showing that computer aided techniques can be used to aid in the attribution process. Thus there needs to be some method of evaluating the results. Thus to test the systems a set of forms are required for which the agents are known to us but not to the classifier. The feature vectors for these forms is the **test set**. The classifier is asked to attributed these and the ratio of the number of incorrect predictions to the number of forms is the test error and represented by ε_T. From the test error we may calculate the probability that the classifier would misclassify a novel form as its true error estimate, which will be abbreviated in this study as error, represented by ε. Often in this study, there is not enough data for a separate test set. Thus a method called **leave one out** (LOO) is used. LOO trains the data on all but one example in the design set and then predicts the value of the left out example. This is repeated many times, each time leaving out and making a prediction on a different example. Using LOO, a set of predictions from the design set can be made and the error may be calculated from the number of misclassifications among the LOO predictions. This study reports not only the error, but also a 95% upper bound which is the a number that we're 95% confident is more than the actual error. Intuitively, this figure simply bounds the expected performance and places a limit on our skepticism of how much the error estimate may disappoint us in the real world.

2.7.2 Structure of Chapters

Before moving to the actual findings of the study, a brief overview of the structure of these chapters will facilitate their reading. Each chapter opens with a brief discussion of the chapter's aims. This is followed by an exposition of the art historical theories or motivations behind the approach. This is followed by a discussion of the methods, usually divided into two parts. Each of the studies are conducted on the basis of a series of experiments. These are described in the first part of this section. This generally discusses the aims of the experiment as well as the overall classifier design. The second part of the section describes the data. This includes a discussion of the candidate agents and a motivation for their inclusion in the experiment, the pre-processing of the data, and a discussion of the feature extractors used. The final section presents and discusses the results.

Chapter 3

Digital Morellian Analysis

This chapter presents the first of the techniques for computer-aided attribution. The technique is based on one of the foundational art historical attribution methodologies, that of Giovanni Morelli. The chapter has a threefold aim. First, the method of Morelli is explained, and secondly it is applied to the corpus of the Princeton Painter in the traditional fashion. The third part, and bulk, of the chapter presents a computer-aided methodology for implementing Morelli's method, and demonstrates it on one of the Morellian features that defines much of the Princeton Painter's corpus.

3.1 Background: The Method of Morelli

The origins of the scientific connoisseurship that is used in the study of vase painting can be traced to the techniques developed by the Italian politician, physician and amateur art historian, Giovanni Morelli (1816-91). Morelli, dissatisfied with the way in which art critics of his day propounded their ideas without any substantiation (beyond the weight of their own renown) developed an 'objective'[1] method of attribution based on an analysis of the way painters rendered minor details such as hands, ears and drapery. The basic premise is that, in his grand compositions, his themes and his choice of palette, an artist exercises conscious choice, which is influenced by a multitude of factors (such as the artist's mood on the day he painted the particular work) that may have varied from day to day. On the other hand, in painting (or incising) the minor details, he recalls a set of preset patterns (which Morelli called *grundformen* and which I will call forms sets) that he has learned through years of repeated practice, and renders these without conscious reflection. In other words, an artist renders a hand or a knee without thinking because he has practised rendering these features so often that they are almost instinctive. Thus they would be rendered consistently from painting to painting.

Morelli did not advance any theoretical justification for his method beyond the obvious, and it has come under some criticism in the past. However, despite claims that Morelli's method cannot be justified theoretically, it is consistent with what we know about the nature of skill acquisition and learning. To explain how, it is necessary to distinguish between two subtypes of long term memory - declarative and non-declarative. The former is associated with concepts, semantics and biography while the latter concerns emotions and skills [Tulvig, 1972]. In particular, the memory associated with

[1] Perhaps 'empirical' is a better term because the system is based on observation. It certainly isn't free from the particular bias of the art historian because they choose the initial categories.

the learning of skills is called procedural memory and can be accessed without conscious effort to the extent that experts using well learned procedures are often unable to articulate the rationale for what they are doing. Furthermore, procedural memory is extremely long term[2] and not subject to radical adjustment, meaning that procedures learned this way remain consistent. Thus the act of painting relies both on procedural and declarative memory. Unifying themes in the painting, about which the artist will usually reflect for a long time before even starting the project, generally use declarative memory. On the other hand, the actual motions of the brush strokes will use procedural memory in such a way as to meet the aims that have been set by reflection. The rendering of minor details is unlikely to require conscious reflection and may well provide a good key to the painter's signature.

Morelli's method is of central importance in the study of black figure. Beazley almost certainly used Morelli's method as the starting point for his attributions. After Beazley, many subsequent studies on individual vase-painters use this technique as a primary tool in the analysis of a painter's style. Typically, scholars make some attempt to define how the painter renders certain minor details, usually anatomical details or subsidiary decoration. It is thus the logical starting point for the development of a computer-aided technique of attribution. Furthermore, the problem of recognising similarity between various form-sets is a shape-recognition task - a very well-researched area of Computer-Science and one for which many algorithms have been developed.

In addition, a computer-aided technique may solve some of the problems inherent in Morelli's method. Morelli's method has been cumbersome to implement. Often, monographs on individual painters include many drawings of minute details from each of the paintings of the artist in question. The process is laborious and also counter-productive to the aims of disseminating findings. First, thorough analyses conducted according to this procedure are only suitable for monographs since they are too large for journals. Secondly, and by corollary, when attributions are published in journals, they are necessarily superficial since they hide the vast amount of work that has already been done. This is a great pity, since monographs are very expensive to produce and hiding much of the work done to make an attribution (such as necessary in journal articles) means that a great deal of the justification for an attribution is never revealed to other scholars. A computer aided-method for Morellian analysis will ultimately solve this problem, since the attributions will be done by relatively unbiased classifiers, dispensing with the need

[2]There is even evidence suggesting that it may not be subject to age related memory loss in the same way that declarative memory is [Churchill et al., 2003]

for the corroborating evidence of multitudes of drawings of forms.

In this chapter it is demonstrated that a classifier may be designed that can carry out a large portion of the Morellian technique automatically. This is demonstrated by training a classifier on forms from 3 different Morellian form-sets, associated with the Princeton Painter, Exekias and Group E respectively. The classifier appears capable of correctly identifying the painter associated with the feature at least 93%[3] of the time. The next section of the chapter comprises a description of the most important Morellian characteristics of the Princeton Painter's style: a definition of his signature, so to speak. This is conducted in more or less the traditional way. The section that follows this describes the design, implementation, and the evaluation procedure for the classifier, and then assesses its performance. Finally, the conclusion considers and interprets the results.

3.2 The Major Princeton Painter Form Sets

The Princeton Painter is particularly well suited to Morellian analysis because he has some very distinctive methods of rendering anatomical details. In particular, he renders most anatomical details with very economic use of incision. For example, his mouths are often rendered with a single incision (table 3.2). This may reflect his leaning towards mass production rather than the production of masterpieces.[4] Whatever the reasons, because they are so distinctive, the Princeton Painter's Morellian features are very good candidates for digital classification. What follows below is a list of the features of the Princeton Painter that are most characteristic of his style. In particular, certain forms-sets, such as the Type A ear, are signature forms as they occur very infrequently outside of the Princeton Painter's corpus. The digital techniques presented in this chapter will be demonstrated on the type B knee (table 3.3: TYPE B).

Throughout this chapter, the term **form** will be used to refer to an instance of a Morellian feature such as a specific knee on a specific painting. The term **form-set** will be used to denote a set of forms that all share properties that make them diagnostic criteria for attribution. For example, the method of rendering a knee with a "C" shape above a short line-segment describes a type of form that is very common in the Princeton Painter's corpus and very uncommon outside of it. One form-set associated with the Princeton Painter, therefore, is the set of all knees that match this description. From here on, the terms form and form-set will be used and the terminology is ex-

[3]Based on the 95% upper bound of the error. The accuracy itself is 97%

[4]Discussed further in the appendix (A.3)

plained more formally in the section concerning the digital implementation of Morelli's method (3.3.2)

The rest of this section proceeds by describing specific form-sets that are associated with the anatomy of the Princeton Painter's male human figures. This is not a complete description of the Princeton Painter's style, but this limitation was necessary for practical reasons (a more thorough description would have been a thesis in itself) and because the male human anatomy provides the most important diagnostic Morellian features in the author's opinion. Each of the key features, ears (3.2.1), mouths (3.2.2), knees (3.2.3), and greaves (3.2.4) is discussed very briefly and the key elements of each form-set described in a table. The table includes the criteria for membership of the set, a list of vases on which forms from this set are rendered, and an example image with the relevant form highlighted. The details of the vases from which the examples are taken appear in footnotes.

3.2.1 Ears

Because the heads of most figures in Attic vase painting are rendered in profile, facial features are not only common, but are rendered from exactly the same angle. This restriction placed on the artist by an iconographic tradition that did not allow for multiple viewing angles means both that ears are rendered on almost all uncovered[5] faces in Attic black-figure, and that they are rendered from the same perspective allowing easy comparison between painters. This makes them an ideal feature for Morellian analysis in vase paintings.

While many of the great vase painters strive for realism or ornate detail in their rendering of anatomical details like these, the Princeton painter appears to strive for economy sometimes at the expense of visual effect (table 3.1).[6] For example, the type A ear (table 3.1: TYPE A) is very clumsily and hastily rendered and the effect is poor. However, a second type that appear in his corpus (table 3.1: TYPE B) is a better compromise between effect and economy, but there is still considerable variation in the way in which this form is rendered. The type C ear (table 3.1: TYPE C) appears to be the most mature. Its effect is good, but because it is a simple shape, it can be rendered with a few incisions and minimal effort. This particular form-set is one of the most distinctive of the Princeton Painter.

[5]for example, by helmets.

[6]Form A: New York 56.171.9 **M10**; Form B: Geneva HR 84; Form C: London B212 **R9**; Form D:Munich 1378 **E1**.

Table 3.1: The Princeton Painter Ears

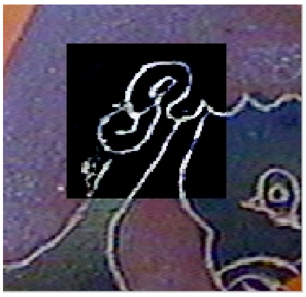
	**TYPE A** A very crude and simplistic form, the ear is incised with a hook or C shape, which may have some further elaboration such as a circle incised inside. This ear looks as though it has been executed in a rush, and not much care has been taken to distinguish the incisions of the ear with those indicating side-burns. **NY 23.160.92; Munich 1378; Princeton 168; Bonn 45**
	**TYPE B** The incision forms a shape akin to the number 8 or the capital B. It is composed of two ovals, one above the other. Either or both ovals may contain further incised details inside. **Munich 1378; Geneva HR 84; Brussels R279**
	**TYPE C: SIGNATURE FORM SET** The ear is composed of two concentric S shapes. This is by far the most common type of ear found in the Princeton Painter's corpus. This form almost never appears outside of the Princeton group. And even within the Princeton group, almost all occurrences are from within the corpus of the Princeton Painter himself. **Forms from this set occur on the vast majority of vases by the Princeton Painter**
	**TYPE D** The shape is similar to the type C, except instead of two concentric S shapes, only a single S is used, typically with a small incised dot or line in each of the cavities of the S-shape. The form is used quite often in a group of vases that I have suggested may be earlier works (A.5.3). The image on the left is interesting in that it appears on a female and it is rendered in reverse.

Table 3.2: Princeton Painter Mouths

3.2.2 Mouths

The Princeton Painter mouth has a standard form: it is indicated with either a single or two very short incisions. This minimalism is not unique to the Princeton Painter but is also apparent in the work of some of his contemporaries, and in particular, the Painter of Berlin 1686 with whom there is a distinct stylistic similarity and with the Swing Painter whose style was close enough to that of the Princeton Painter that Beazley suggested the latter was the pupil of the former.[7] In addition to the single line mouth, the Princeton Painter occasionally added a moustache, which is rendered with two parallel arcs extending from just below the nose, around the mouth and terminating on the beard, for example in the image on the right of table 3.2.2.[8]

3.2.3 Knees

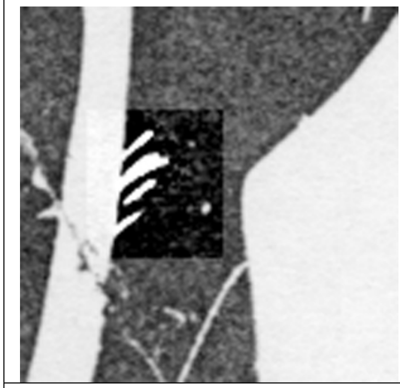
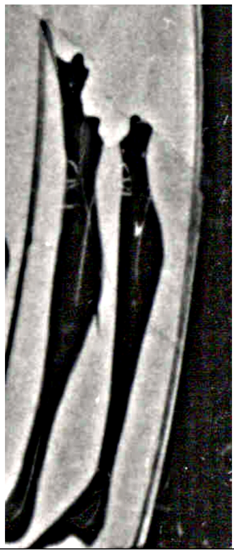
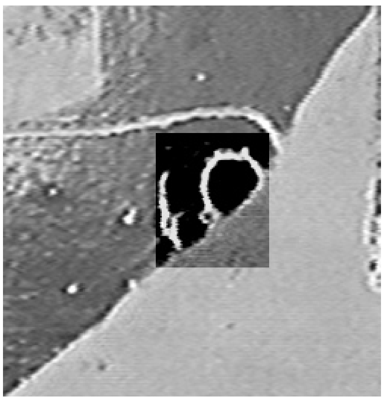
The Princeton Painter's knees (table 3.3)[9] show equal economy to his other form-sets. The simplest is just a couple of incised lines indicating the folds(table 3.3: TYPE A). A better compromise between economy and effect is the type B knee (table 3.3: TYPE B) but the best of these is the type C knee (table 3.3: TYPE C). While this is an admittedly subjective judgment, the Princeton Painter certainly seemed quite proud of this form-set since the

[7] *ABV* 304, an association denied by [Böhr, 1982, p.56]

[8] Left: Cambridge GR 1.1889 **M7**; Right: London B212 **R9**; *Add*[2] 78.

[9] Form A: New York 53.11.1 **E3**; Form B Basel BS 427 **EM1**; Form C: Cambridge GR 1.1889 **M7**.

Table 3.3: Princeton Painter Knees

	**TYPE A** Three, four or five lines in parallel, generally at an angle to the front of the shin. This is quite an uncommon method of rendering knees and we have conjectured that it is a very early approach. **New York 53.11.1; Princeton Y168; Munich 1378; Basel Bs 427; Bonn 45; Madrid 10925**
	**TYPE B** The folds of the knee are rendered by a series of roughly horizontal lines that typically at various angles to each other. In some cases, the two top lines form a < shape. **Princeton 168; Munich 1378; New York 53.11.1; Basel BS 427; Bonn 45**
	**TYPE C: SIGNATURE FORM SET** The folds of the knee are indicated by a semicircle above an arc parallel to the bottom of the semicircle. The open side of the semicircle faces the shin. Forms from this set are very common in the Princeton Painter's corpus. **Forms from this set occur on the vast majority of paintings, with notable exceptions being Munich 1378 and Princeton 168**

vast majority of his vases have examples from it. In fact, the painter appeared to be so fond of this form that he renders the folds of the knee over the greaves. The popularity of this type of knee in the Princeton Painter's corpus and the scarcity of it outside of his corpus make it an ideal subject for digital Morellian analysis, and it is the example used in this study.

3.2.4 Greaves

Greaves are armour leg-guards used in combat and are attested from as early as the 14th century BCE in Greece[Fortenberry, 1991] and were described in Epic poetry.[10] Greaves could be made from a number of materials, but were most commonly made of cloth, leather or metal, and they could cover the leg up to either the knee or the thigh. In vase-painting, greaves usually extend only to cover the shins and the knee, although occasionally thigh-guards are also clearly evident. On Greek vases, greaves are worn both by hoplite and heroes. In some cases, Attic vase painters decorate greaves with designs that may either be meant to mimic the creases in the metal, or elaborate designs. However, most artists do not decorate their greaves. Here the Princeton Painter's technique is somewhat different since not only does he usually decorate the greaves regardless of whether the wearer is hoplite or hero, but he also often does this with an unusual motif whereby an incised arc decorates the inside greave (table 3.4: type A)[11] and a spiral decorates the outside (table 3.4: type B). Thus, if a sherd or pot has images of soldiers, the manner in which the greaves are decorated may well be used as a method for attributing to the Princeton Painter.

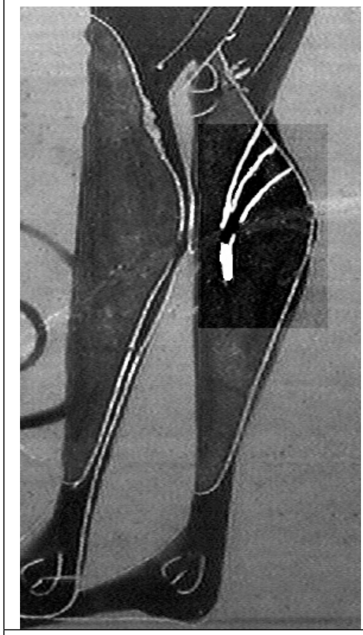
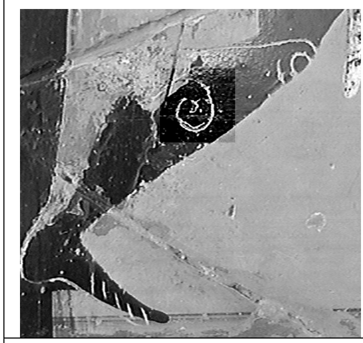
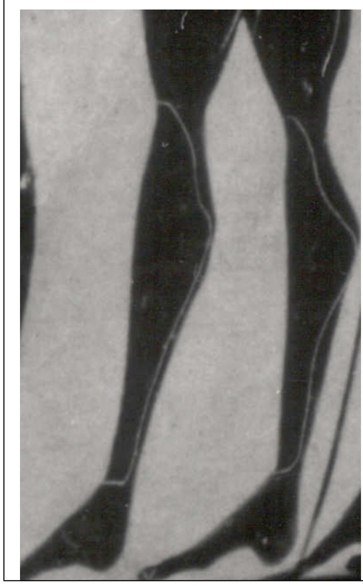
3.3 Digital Morellian Analysis

Morelli's technique is an obvious starting point for a digital style recognition system. The most successful of Durham [1996]'s techniques involved the shape description of images of calves painted by the Antimines painter and the Swing Painter. Although Durham does not explicitly mention Morelli, his explanation for the success of the system is entirely consistent with Morelli's theory. Durham alludes to this by recognising that the Antimines painter renders his calves with an arc to indicate the musculature whereas the Swing Painter does not. This is very typical of the Morellian approach to attribution based on this sort of minute anatomical details. The present study builds

[10] For example, Homer *Iliad* 3.330-1 describes Paris' greaves in considerable detail.

[11] Form A: London B212 **R9**; Form B: Cambridge GR 1.1889 **M7**; Form C: Basel 427 **EM1**.

Table 3.4: Princeton Painter Greaves

	**TYPE A: SIGNATURE FORM SET** An incised set of arcs start slightly apart at the shin and extend downwards converging to a point half-way down the lower-leg and mid-way between the shin and the calf. The type is very rare outside the corpus of the Princeton Painter.
	**TYPE B: SIGNATURE FORM SET** An incised spiral, similar to that appearing on breastplates, appears on the lower leg, probably indicating decorations on the metal. This type is very rare outside the corpus of the Princeton Painter.
	**TYPE C** A plain greave: either undecorated or painted red or white.

on Durham's approach in a two fundamental ways. First, it is up-scaled to a multi-class problem and secondly, the classification system presented is empirically evaluated.

One of the key features of the method proposed in this chapter is the use of artificial data. These are Morellian forms drawn by the art historian (in this case, the author) in the manner of the respective painters so as to create an artificial design set. The technique is inspired by the use of virtual samples in a number of problem domains. Virtual samples are usually samples from the original design set that have been deformed such a way as to approximate variances in the population, for example by using CAD systems to change the orientation of objects in the design set to create new design set samples. Thus, the virtual samples increase the size of the design set often allowing more robust classifiers to be built. These have been shown to improve the performance of certain systems in a number of problem domains including face recognition [Bin et al., 2003] and the detection of holes in flanges [Kuhl et al., 2004] and in the case of the latter, the design set was made entirely of virtual samples.

Additional motivation for the use of artificial design data may be derived from the bias and variance decomposition (2.3.1) and the No Free Lunch theorem (2.4.1). The No Free Lunch theorem suggests that learning cannot take place without an inductive bias. With simple classifiers like the Naïve Bayes Classifier, k-nn and LDA, the bias is often in the form of unrealistic but acceptable assumptions about the nature of the data (in these examples, respectively, that the variables are independent, objects from the same class cluster close together in the feature space, and the class distributions are homoskedastic and normal). Bias that encodes prior knowledge is even better. In this sense, artificial design data allow the art historian directly to encode their domain specific prior knowledge of these form sets by means of visual examples. Throughout this chapter the term 'virtual sample' and 'artificial data' are used interchangeably, and likewise 'real samples' and 'real data'. It should be pointed out, from the outset, however, that virtual samples are only used as part of the design set and never as part of the test set.

3.3.1 Aims

The aim of this chapter is threefold. The primary aim is to demonstrate empirically that a classifier may be trained to distinguish between form-sets associated with different vase-painters. However, an additional aim is to determine whether prototypes of these forms sets rendered by art historians in the manner of a particular agent may be used to train the classifier to recognise that painter's form-sets, and if so to investigate the effect of these artificial data on the performance of the classifier. Finally, the performance of a majority vote ensemble is compared with the performance of the individual base classifiers to determine if this ensemble effectively improves the overall performance of the system. Four experiments are conducted to meet these aims, all of which conform in some way to the classifier design described in figure 3.1 and is described in more detail in section 3.4.

3.3.2 Formally defining Morelli's method

Morelli's method may be expressed in terms of the vocabulary established in Chapter 2. The actual physical feature - i.e the physical part of the painting or the digital image is a form. For many painters there are specific ways in which these forms are expressed and these may be described by an art historian or by a probability distribution over a feature space that describes the form appropriately. With each of these distributions is associated a form set. Different form sets are associated with different painters, and the task of determining to which painter a physical feature found on an unsigned painting, may be described as a two-stage process. First, the process involves determining to which form set Γ the form γ belongs and then deciding which form sets are associated with which agents. In most cases, there will probably not be a bijective map between the set of agents and the set of forms but rather a joint distribution over the set of agents and form-sets. Thus, to achieve the best results probably requires the choice of form-sets that almost uniquely define the painter.

This may be illustrated with some examples. The Morellian features described in the previous sections are all form-sets associated with the Princeton Painter. Any set of greaves (specific instances of which are individual forms), for example, will be of type A, B, C or none of these three. Thus we may define for greaves four form-sets associated with the Princeton Painter. It is quite possible that a set of greaves by some other agent may also match the description of these forms. Therefore, in a real-world application, one would also like to know what a greave being of type A implies about the agent who rendered it. For example, we would like also to be able to answer the question of whether finding a Princeton painter type A greave on an unsigned painting should imply a 50% or 70% likelihood of the Princeton Painter having rendered the form. For this study, the task has been simplified to demonstrate just the essential concept and it is assumed that classification to a form set is equivalent to attribution to an agent.

3.4 Classifier Design

The first stage of the design uses pilot data to select appropriate feature extractors (described in more detail in section 3.4.7.1). This is achieved by testing a number of feature-extractors (3.4.7) using LDA on a pilot set composed of artificial forms drawn from form-sets associated with a variety of vase-painters. Then four different experiments are carried out in order to address the aims stated above. They are summarised here and described in detail below. The first experiment aims to establish baseline error rates for LDA, QDA and 1-nearest neighbour (1-nn) trained only on 'real' data. Here, the error rate is estimated using the leave-one-out (loo) estimate. This will serve as a baseline measure for comparison with the virtual estimates and with the majority vote ensemble. The three classifiers are here used to determine the effect of the virtual samples and of the ensemble method, and not to compare the relative performance between LDA, QDA and 1-nn. The second experiment examines whether virtual samples may be used as a design set on their own, and whether using virtual samples for validation improves the performance of the classifiers used in experiment 1. The third experiment investigates the effect of artificial data used to augment a design set composed of real data. Here the same feature spaces are used as in experiment 1. However the experiment is repeated with different numbers of artificial samples added to the training set in order to determine the performance of the respective classifiers as a function of the number of artificial samples added to the design set. The final experiment estimates the performance of a majority vote ensemble of all classifiers used in experiment 1, and compares this with the baseline experiments of the individual base classifiers. In all cases the test sets were composed exclusively of real samples.

3.4.1 Experiment 1

For each of the 5 feature spaces selected from the pilot process, 3 classification rules are constructed using LDA, QDA and 1-nn respectively, trained on the real data prepared as described in 3.4.6.3. This results in 15 classification rules. Each of the feature spaces is of high dimensionality so PCA is applied and the first N components that account for more than 90% of the variance are retained in the case of LDA and 1-nn and 70% in the case of QDA (**PCA90/70**). Performance is measured using LOO on the real samples only, and the posterior distribution of the true error is estimated by assuming a uniform distribution over [0,1] for the prior probability with a binomial likelihood resulting in a beta-distributed error estimate, as described in 2.5.1. The error rate and the 95% upper bound are reported. This experiment

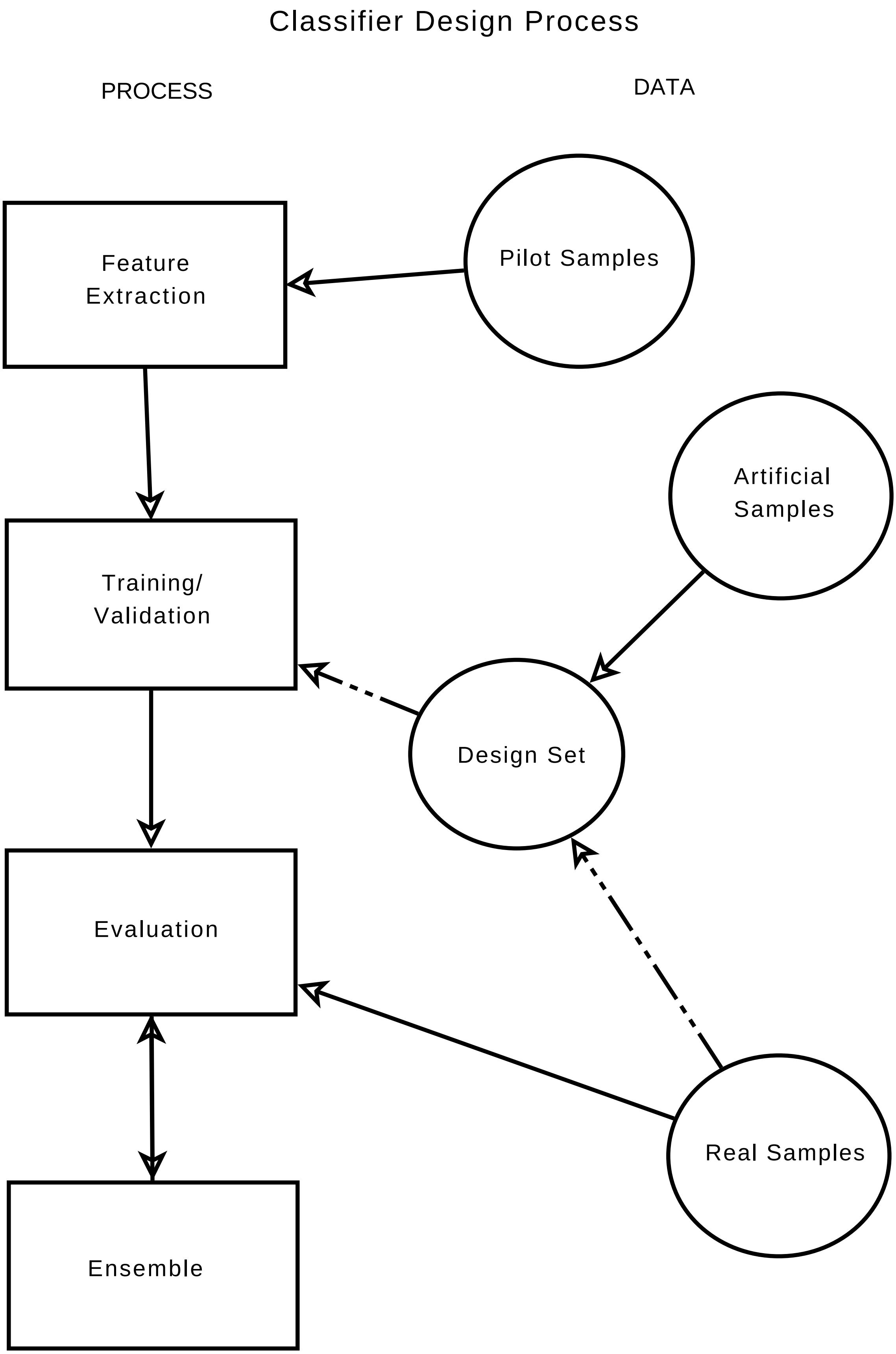


Figure 3.1: conceptual design of classifier.

serves as a baseline measure of the system's performance against which the other experiments will be judged.

3.4.2 Experiment 2

The second experiment evaluates the virtual set as a design set or validation set in its own right. The aim is ultimately to determine both whether the design set may be composed entirely of artificial samples and whether using the virtual samples to select the optimal number of components leads to increased performance over using PCA90/70 for this task. To achieve this three sets of classifiers are constructed. First, a design set is constructed using only the artificial data as described in 3.4.6.4. Then each of the 5 feature extractors is used to train 3 classifiers: LDA, QDA and 1-nn on the artificial design data, resulting in 15 classification rules. Again, PCA90/70 is used to reduce the dimensionality of the feature-space as in experiment 1. Secondly, this is repeated but instead of using PCA90/70, the optimal number of principal components is determined empirically from the design set using gaussian bolstered leave-one-out (BLOO) error estimation (2.5.2). Third, the experiment is performed with a design set comprised only of real samples, but where the optimal number of components is determined by applying BLOO to the virtual set. This allows us to determine whether there is any significant difference between the virtual design set and the real design set, and also whether using 90% of the PCA components is significantly worse than using the optimal components.

On the basis of these sets of classification rules the following pairwise comparisons are made:

i) the performances of the base classifiers trained on the virtual design set versus the baseline measurements (i.e. trained on the real data) using PCA90/70;

ii) the performance of the base classifiers trained on the real design set versus their respective performance when trained on the virtual set, using the virtual set to select the optimal components in both cases; AND

iii) the performance of the base classifiers using PCA90/70 versus their respective performance using the virtual set to select the optimal components.

In all cases, the error and 95% upper bound is determined by testing on the real data only. Where the design set is composed of real data (i.e. in ii and iii), LOO is used, otherwise (in i) the holdout method is used. To determine whether the performance differences in each of these 3 pairwise comparisons are significant, two tests are used. The Wilcoxon T test is used to compare the two methods across all 15 classification rules (i.e. QDA, LDA and 1-nn each trained on 5 different feature sets), while the McNemar test compares each of the 15 classification rules individually with their counterpart in the pairwise comparison. For example, McNemar's test may compare LDA trained on feature space 1 using PCA90/70 with LDA trained on feature space 1 using validation. The tests are described in more detail in 3.4.5

3.4.3 Experiment 3

The aim of this experiment is to determine the effect of augmenting a real design set with artificial data. The issue is motivated by the belief that increasing the size of the design set will decrease the variance of the classifiers and that consequently, flexible classifiers that would usually have overfit to the small design set may start to improve as the variance drops. In order to test this the same feature sets and classifiers are used as those in the first two experiments. The initial design set used is the real dataset, but the experiment is run 20 times. In each successive run of the experiment, one extra sample per class is added to the design set, sampled without replacement from the set of virtual samples. This entire process is repeated 20 times for each feature set, each time with a different set of virtual samples used to augment the design set, and the average for each number of added samples is recorded. The effect of the increase in size of the design set is reported and compared with the baseline from experiment 1 and the artificial training set in Experiment 2. The results are evaluated qualitatively by examining the graphs of error rate on the real data (using LOO) versus number of added virtual samples in order to determine any trends.

3.4.4 Experiment 4

Experiment 4 evaluates the majority vote ensemble to improve the performance of the base classifiers. The idea is that instead of choosing between these classifiers, the classifiers are used in combination, as described in 2.6. To achieve this, using each of the five feature extractors, LDA, QDA and 1-nn are trained on a design set composed of virtual and real data - for a total of 15 classification rules. The classifiers each make their predictions on the real data using LOO. The 15 sets of predictions are combined using a majority vote ensemble. The result is compared with each of the classifiers from experiment 1. The significance of these differences is determined using McNemar's' test.

3.4.5 Comparisons Between Classifiers

The experiments above call for comparisons between classifiers. The issue of how to perform such comparisons is both intricate and difficult to perform, particularly when there is a small number of samples from which to draw conclusions. There are two major issues that should be addressed when making such comparisons, and these are dealt with in turn. The first is dealing with the variance in the estimate resulting from the data and the second is dealing with the variance due to multiple simultaneous hypotheses testing.

In the first instance, the data used in the design of a classifier leads to two sources of variance that affect how we interpret a point estimate of a classifier's performance as a measure of its ability to perform in the real world. The first is that the classifier is tested on a small test set, so there is always some error inherent in extrapolating to the population at large. The second is that it is trained on a design set which is itself an approximation of the population and this itself is a source of variance. Since both of these are finite, ideally they should both be accounted for in making such comparisons between classifiers. However, few studies actually do take into account the internal variability of the design set as there are a number of limitations to many of the methods designed to take this into account (for example, most of these tests are computationally intensive). More importantly, for the experiments in this study, these tests are inappropriate for two reasons. First complications arise from the fact that the virtual and real sets are not necessarily drawn from the same distributions and secondly, this study also makes direct comparisons between classifiers with different size design sets. Furthermore, McNemar's test (which does not take into account variability in the design set) has been proven in simulations on both artificial and real-world data to be a good predictor of difference in performance between classifiers and outperforms[12] most methods that do take into account the variability of the design set [Dietterich, 1998].

The exposition of McNemar's signed rank test below follows Dietterich [1998]. Let c1 and c2 be two classifiers tested on a set S, and let $\Delta\varepsilon_1$ be the number of items in S incorrectly classified by c1 but not by c2 and $\Delta\varepsilon_2$ be

[12] In terms of both accuracy and power.

the converse. Then the statistic

$$\frac{(|\Delta\varepsilon_1 - \Delta\varepsilon_2| - 1)^2}{(\Delta\varepsilon_1 + \Delta\varepsilon_2)} \tag{3.1}$$

is distributed according to χ^2 with 1 degree of freedom under the null hypothesis. Therefore a figure of 3.815 allows us to reject the null hypothesis at $\alpha = 0.05$. When the denominator is small (say < 10) the statistic is not well approximated by the χ^2 distribution and for the purposes of this study will be rejected. The McNemar test is used in this chapter to compare the classifiers trained on virtual samples with those trained on real samples, to compare PCA90/70 with using the virtual set to select the optimal number of components, and to test the difference between a majority vote ensemble and the individual base classifiers.

When two classifiers are trained on multiple feature spaces to produce multiple decision rules, their overall performance across all the feature spaces may evaluated using a single statistic, the Wilcoxon signed rank or T-test. The intuitive approach may be to simply compare their respective mean performances over all the feature spaces, but the mean is too sensitive to outliers. A more robust test, and the standard non-parametric test[13] is the Wilcoxon T-test [Wilcoxon, 1945]. Given two vectors X and Y, the T-test assesses the hypothesis that pairwise differences between the corresponding elements of X and Y are drawn from a distribution with median 0. Intuitively, if X and Y represent the error estimates of two different classifiers and in which each element of these vectors represents a different feature-space, then if the overall performance of the two classifiers across all of these feature spaces is the same, the median of their differences should be 0. This test is recommended in such circumstances by Dietterich [1998].

The test works as follows: Let c1 and c2 be two classifiers and let there be N trials, each using a different design set composed of the same objects, but represented in different feature spaces. Let d_n be the absolute difference between the performances of c1 and c2 using design set n where $1 < n < N$. These differences are ranked from largest to smallest. Let r_n be the rank of d_n and let $S_n = 1$ if c1 outperformed c2 on set n, and 2 if c2 outperformed c1 for the same set. Then

$$T = \min\,[\sum_{n=1}^{N} r_n \delta(S-1), \sum_{n=1}^{N} r_n \delta(S-2)] \tag{3.2}$$

where $\delta(.)$ is the Kronecker delta function. Critical T values for different N can be found by examining an appropriate table of scores, supplied in

[13] ANOVA is usually preferred when the data are normally distributed

most standard statistics texts including Siegel [1956] which popularised the method. In this study, a p score will be reported.

A number of the results reported in these experiments make direct comparison between many different classifiers. This amounts to a multiple hypothesis test and in such cases the level of significance needs to be adjusted to take this into account. Multiple hypothesis testing is the subject of much intense debate and the philosophical as well as the practical approaches to solving the problem differ widely, not only between Bayesian and Frequentist camps but also within them. The problem may be stated quite simply. When multiple tests are conducted simultaneously, if the level of significance for each hypothesis is set at $\alpha = 0.5$ then the probability of one of the null-hypotheses being falsely rejected is proportional to the number of simultaneous tests. We refer to such a false rejection as a **false positive**. Assuming there are N tests and that α is the experiment-wide significance level and α_n is the significance level for hypothesis $n \in \{1..N\}$ then $\alpha = 1 - (1 - \alpha_n)^N$ if we assume independence between the respective tests. Regardless of independence, the following holds: $\alpha < N\alpha_n$. Intuitively this means that no matter what level of significance you set, if you repeat the experiment many times, you are likely to get false positives. For example, if you are looking for to test the effectiveness of 20 different treatments, setting $\alpha = 0.05$ you would expect by chance that at least one of these treatments will appear to be effective even if it isn't. Therefore one is not on firm ground if, having found one treatment that appears to be significant, one reports that treatment to be significant at the $\alpha = 0.05$ level.

One solution to the problem is to reduce the statistical power of the tests such that one guarantees that the α_ns are all at least α by setting $\alpha_n = \frac{\alpha}{N}$. This is the famous Bonferroni correction which for most practical purposes is overly conservative and has the potential, when N is large, to reduce the statistical power to such a degree that the type II error rate is unacceptably high - indeed in some cases where the system is incapable, in practice, of producing true positives (such as in microarray gene expression tests where or fMRI scans when very large numbers of simultaneous tests are conducted). However, in the cases where N is small and the individual hypotheses are of interest in themselves, Bonferroni's method is acceptable. For this reason, this correction is employed in this chapter for comparisons between classifiers, although it is used cautiously.[14].

[14]The more complicated Bonferroni-Holm stepdown procedure is used in Chapter 5.

3.4.6 The Data

3.4.6.1 The form Sets

The principal is tested on a form set associated with the Princeton Painter, the type C knee, and two knee form-sets associated with Group E and Exekias respectively. While Group E and Exekian form sets often have some overlap, the two particular form sets chosen to represent these two agents are quite common in their respective corpora. An example of the Exekian form-set is indicated in figure 3.2(a). The main curve has a pronounced bulge towards the bottom, and in the cavity to the right is a small incised curve. In addition, there are also occasionally other incisions either underneath the main curve or in the cavity. An example of the Group E form-set is indicated in figure 3.2(c) and may be describe loosely as a curved L-shape, but with the vertical component leaning somewhat to the right. In the cavity there is usually at least one small incision and sometimes more.

3.4.6.2 The Agents

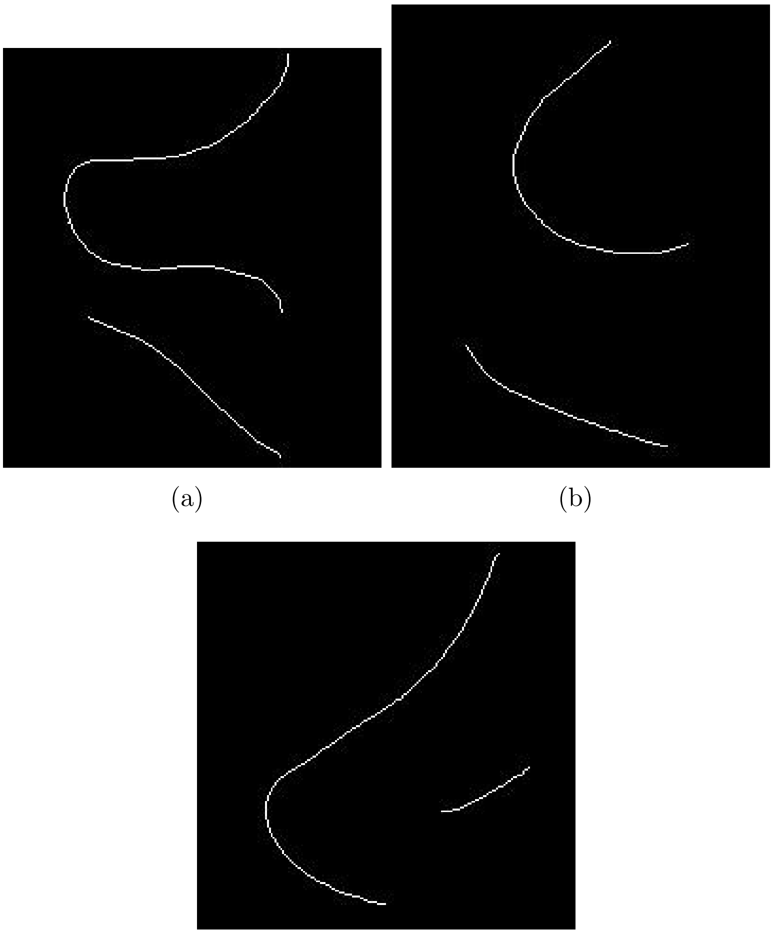
Group E and Exekias are very closely related although the latter's earlier extant works are somewhat later than those of the former. Rough dates for Group E are between 560 and 540 and the Princeton Painter between 545 and 530. Beazley describes Group E as

> "...the soil from which the art of Exekias springs, the tradition which on his way from fine craftsman to true artist he absorbs and transcends."(*ABV* 133.)

In fact, Group E are named after Exekias, despite being earlier. Because of the close relationship between these painters, it is unsurprising that the forms they use are quite similar. This, however, provides a good test of the system since, to the untrained eye, some of the Exekian knees in the test set are very close to some of those of Group E and vice verse. On the other hand, the Princeton Painter, although in some way close to Group E, is not nearly as close as Exekias. The form set chosen is quite possible the most diagnostic of the Princeton Painter's style.

3.4.6.3 Real Data: Image Processing

Because the photographs were taken under many different conditions which were beyond the control of the experimenter, the real data could not be processed by automatic techniques alone. However, the method used is consistent from sample to sample. 20 forms are sampled from each of the form



(a) (b) (c)

Figure 3.2: . Examples of processed images of the incisions indicating knees by (a) Exekias (b) The Princeton Painter and (c) Group E.

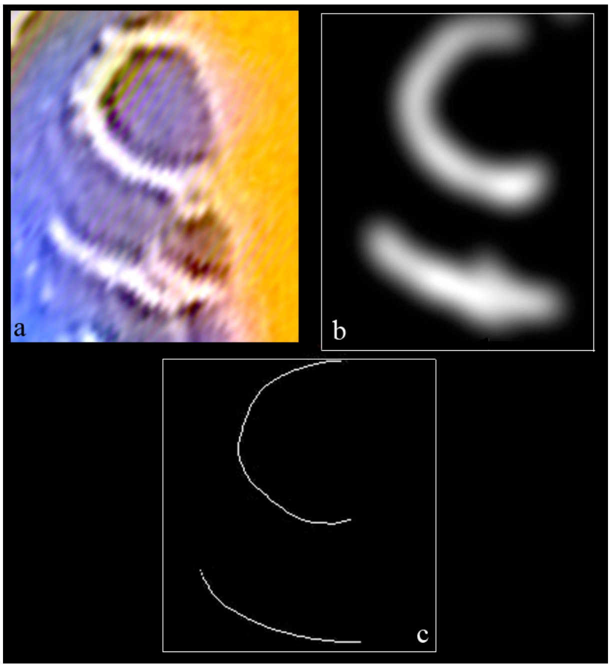


Figure 3.3: 3 Stages in the processing of the incisions of the knee. (a) is a close detail from the knee of a warrior on Cambridge GR 1.1889 (b) the incision extracted and blurred and (c) the final skeletonised image.

sets and the incisions from these images are separated from the rest of the image using the selection mask function in Photoshop. These are then placed on a black background and scaled so that the figure is 200 pixels high. The images are then blurred and contrast stretched so that the image is black except for the area around the incision which is light gray, and where the intensity is generally greatest in the centre of the incision. The figures are printed out and a curve is traced through the middle of the blurred area, using the brightness as a guide. Three stages of the process are illustrated in figure 3.3 These are re-scanned and scaled so that the figures are 200 pixels high, and the figures are skeletonised using the medial axis transform with pruning.

3.4.6.4 Virtual Samples

Unlike true samples, which are 'sampled' directly from the real world objects, virtual samples are generated artificially in such a manner as to resemble real samples. Virtual samples have been used to considerable success in solving the small sample problem in face recognition, particularly using eigenface methods, in which often the number of samples is significantly smaller than the number of variables measured. While in these cases, virtual samples are created by an algorithm that deforms real samples in a variety of ways so as to simulate real-world noise, in this present study the virtual samples are created by a human as they imagine to be representative of the respective form-set. In effect the term 'virtual sample' is a complete misnomer since, having been created by a human, they are not virtual, and they are not sampled. I keep the misnomer for historic purposes since the idea is the same.

Since the virtual samples are not sampled from the same source as the real samples, they have the potential to harm the performance of a classifier if used indiscriminately. There are two reasons virtual samples may be useful for archaeologists. The first concerns cases where the archaeologist has no photographs of sufficient quality that allow for a design set composed of real samples. In such a case, intuition would suggest that virtual samples are better than no samples at all. The second case is where some real samples are available and virtual samples are used simply to increase the size of the dataset. Considerable care should be taken in the second example. The goal of experiment 1 is to determine whether a system could be trained entirely by virtual samples. Experiment 2 determines whether, given an existing design set of real samples, virtual samples may be used to enhance the predictive power of the classifier.

The manner in which the virtual samples are created is straightforward. The art historian, in this case the author, draws the forms in red pen on white paper. The images are scanned and rescaled so that the figure from top to bottom is 200 pixels high, and then thresholded leaving the formerly red pixels as 'on' and the white as 'off'. In both the virtual and real samples, the images were drawn to be as similar in size as possible and were scanned at the same resolution to ensure that artefacts from rescaling, if any, would uniformly effect the dataset and not introduce significant sampling bias. Such artefacts have the potential to increase the noise in the system and adversely effect performance.

3.4.7 Feature Extraction

The Ugly Duckling theorem [Watanabi, 1969](2.3.3) implies that there is no universally optimal set of features, but instead that features can only be described as good or bad in relation to the specific problem to be solved. As explained in 2.1.2, the two usual approaches to solving the problem are to select shape descriptors that are invariant to artefacts (i.e. features that are not relevant to the classification task) and to survey the literature for descriptors that perform well on similar classification tasks. Unfortunately, for the particular problem it is not clear exactly what would constitute irrelevant shape information, apart from topological features (for example, these shapes almost all have the same Euler number) and scale. Instead, therefore, the literature has been surveyed for shape descriptors and these have been selected based on their performance in a pilot round, the exact nature of which is beyond the scope of this dissertation.

However, again, the particular problem under consideration here has no exact analogues in the literature although it shares some similarity with optical character recognition and offline signature verification as both these are interested in extracting shape details from binary images, often skeletonised, images. One major difference, particularly with OCR, is that context is not important in this study as the forms are measured in isolation from their context. In OCR on the other hand, the letters are never only considered in isolation, but as part of the word as well - often using time series and state estimation techniques like Hidden Markov Models. However, some of the feature extractors used in OCR, particularly in earlier studies, offer some useful parallels. The literature on signature verification also has some overlap with the present study. In both OCR and offline signature verification feature extraction, 2D shape descriptors are often used to extract basic shape information. A survey of some of the more popular methods reveal a number of different approaches to shape description. To determine which of these should be used in this study, a number of candidate techniques were tested on a pilot set.

3.4.7.1 The Pilot Studies

The technique used to tackle feature selection in this study is to apply a small group of feature extractors to the selected processed images and to apply PCA to these for dimensionality reduction. Selecting an appropriate set of feature extractors to start with required some sense of what would be likely to work and what wouldn't. For this purpose, a number of pilot tests were conducted in which a large variety of feature extractors were evaluated intuitively by the author. The extractors were tested on various sets of line-drawings rendered in the manner of different painters, including the Princeton Painter, but excluding group E and Exekias.

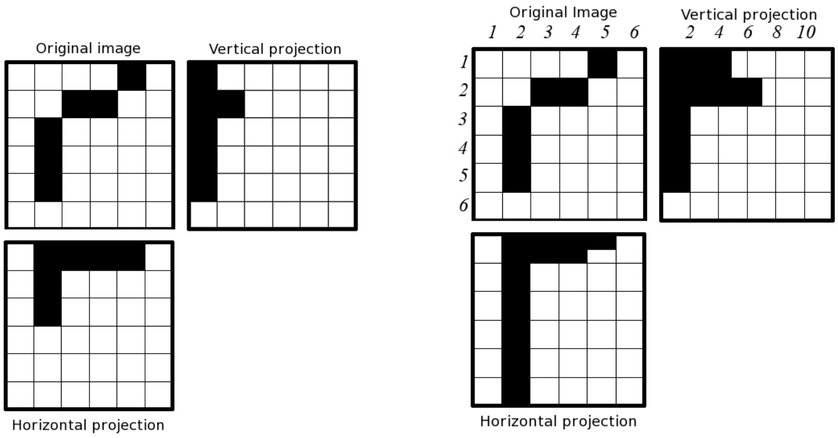
The selection of feature sets is conducted as follows. For each of the forms in each of the form-sets in the pilot data, a feature vector is constructed using a variety of different feature extractors sampled from the literature, particularly in OCR and offline signature recognition, including descriptors modified by the author. These are all compressed using PCA and the first three components plotted on a 3D scatter graph. The feature extractors are evaluated intuitively by the author on the criteria that they scatter in such a way that they look separable. Then LDA is trained and tested on the fewest principal components that account for 90% of the data and evaluated using the leave-one-out estimator. Using a combination of an intuitive appraisal of the scatter plots and the final error estimate on the test set, the following feature extractors proved the best and were retained as the basis of the experiments.

Of all the feature extractors used, those that appeared most appropriate from the pilot test were horizontal and vertical projections, the H-transform, curve signatures and quadrant statistics. These extractors are explained in detail using the following terminology. A grayscale image is represented by, $\mathbf{I}$, which is a $y \times x$ matrix of graybeard values between 0 and 1, $\mathbf{B}$ represents the black and white image (in this case, a pruned skeleton) and is a $y \times x$ matrix of binary values, and $\mathbf{i}$ represents the unit column vector.

3.4.7.2 Vertical and Horizontal Projections

The vertical and horizontal projections are among the simplest of transforms that may be performed on skeletonised images, yet they have proven very effective on a number of optical character recognition problems [Sonka et al., 2007, p.256]. In its simplest form, the vertical projection is a count of the 'on' pixels in each row of a binary image and the horizontal projection is the pixel count of the columns. This is achieved simply by $\mathbf{Bi}$ for the horizontal projection, where the dimensionality of $\mathbf{i}$ is the width of $\mathbf{B}$ and similarly the vertical projection is $\mathbf{B}^T\mathbf{i}$ where the dimensionality of $\mathbf{i}$ is the height. This is illustrated in figure 3.4(a) where the first row has only 1 on pixel so the vertical projection for that row is 1, whereas column 2 has 3 'on' pixels resulting in a horizontal projection of 3 for that column.

This study does not use this simple version of the projection. The reason is that while the vertical and horizontal projection are sensitive to gradients that are particular steep and are revealing if the skeleton has lines that cross or if lines overlap, in images where there are few such features the projections are unrevealing. There are a number of ways to increase the discriminative information in projections. One is to convolve the image with a gaussian kernel and apply the transform to the resulting grayscale image. For this study, however, a novel method is employed whereby each pixel is weighted by its respective distance along the axis of measurement. Thus we introduce the column vectors $\mathbf{d}_h = [1, 2 \ldots X]$, $\mathbf{d}_v = [1, 2 \ldots Y]$ where X and Y are the width and height of $\mathbf{B}$ respectively. Then the weighted transforms are $\mathbf{Bd}_h$ and $\mathbf{B}^T\mathbf{d}_v$ respectively. In figure 3.4(b), for example, the only pixel



(a) Simple Projections (b) Variant

Figure 3.4: Horizontal and Vertical Projections. (a) illustrates the simple or standard projection, and the (b) illustrates the modified projection used in this study. For this example, black pixels are considered 'on'.

in the first row of the original image is 5 pixels from the left and so the vertical projection for that row is 5. On the other hand, the second row has two on pixels at positions 3 and 4 respectively, giving a projection of 7, and finally column 2 of the original has pixels 3 4 and 5 'on' giving a horizontal projection of 12 for that row. This is illustrated in figure 3.5(a).

3.4.7.3 The $\mathcal{H}$-Transform

Inspired by Tabbone et al. [2006], who defined a 1-D descriptor on the Radon transform, we define on the Hough transform an analogous 1-D descriptor called the $\mathcal{H}$-transform. Before explaining the transform, a description of the Hough transform is required. The Hough transform is a term used to describe a variety of shape description algorithms. In this study the term is used specifically of the line detection algorithm presented by Hough and implemented by Duda and Hart [1972]. The generalisation of the technique to include arbitrary curves [Duda and Hart, 1972, Ballard, 1981] will be referred to as the generalised Hough transform (GHT) and although similar to the method we describe as the curve signature, will not be used in this study.[15] The basic idea behind the Hough transform is as follows. All possible lines

[15]The GHT is a very useful shape descriptor, but it relies on the parameterisation of a prototype shape against which other examples are compared. This study is tested on only 3 types of form-sets, but it is expected to be general enough to distinguish between

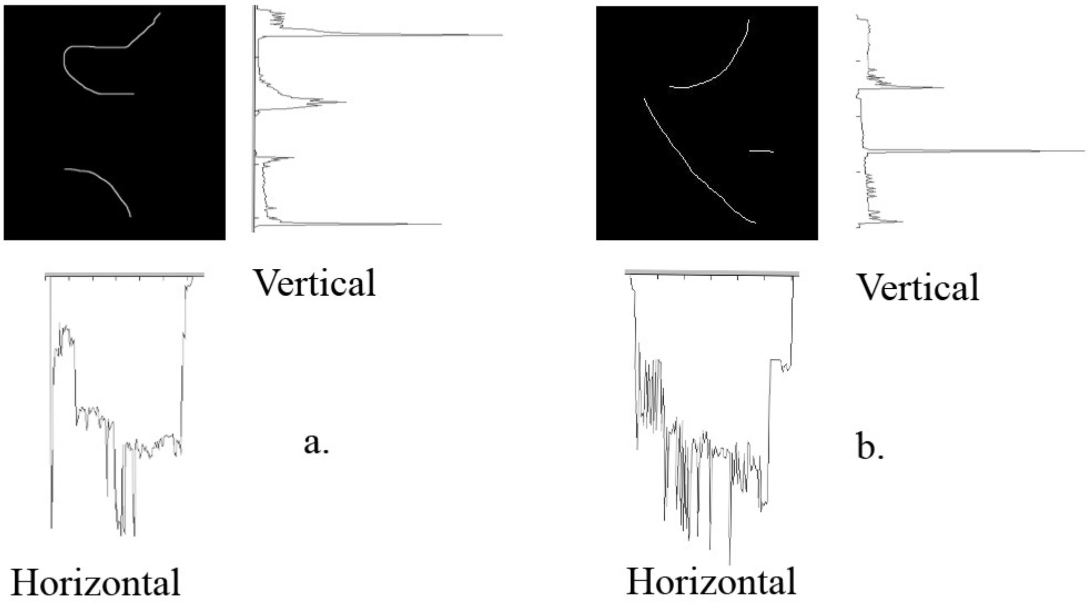


Figure 3.5: Horizontal and Vertical Projections of real knees. (a) A knee by Exekias (b) A knee by Group E. In these examples, white pixels are 'on'.

$\rho = x\cos\theta + y\sin\theta$ in an image may be described by two parameters ρ and θ. Thus in parameter space, every line is described by a unique point. If the parameter space is divided into $P \times \Theta$ bins then a $P \times \Theta$ matrix $\mathbf{M}$, called the accumulator array, may be defined such that the entry $\mathbf{M}_{ij}$ contains the number of 'on' pixels in a binary image that lie on a line whose parameters are in the bin $\rho = i, \theta = j$. Lines in the image may be detected by choosing accumulator values above a specific level. These will be the lines with the greatest number of pixels. The hough transform is robust in the presence of noise and partial occlusion of lines. Rotation of the image results in a translation of the hough transform and scale invariance may be achieved by dividing the hough transform by its mean. There are a number of methods of implementing the hough transform fast, and this study uses the MATLAB implementation of the transform.

The $\mathcal{H}$-Transform, in turn, assigns to every angle in the accumulator a measure of amplitude.

---

a wide variety of form-sets. Therefore the use of a shape prototype is inappropriate.

$$H_a(\theta) = \frac{1}{P}\sum_{\rho=0}^{P} M \tag{3.3}$$

$$\tag{3.4}$$

This measure is simply the mean of the accumulator matrix with respect to ρ[16] as a function of θ, which we will refer to as mean_ρ. Using this terminology, the transform may be written $\mathcal{H}_\dashv(\theta) = \text{mean}_\rho(M)$. This terminology will be used again in chapter 5 and expanded on in the relevant section (5.4.1). Intuitively, the H_a is a measure of the angles which contain most linear elements. It is expected that the $\mathcal{H}$ transform will be particularly good at distinguishing between forms that have many linear features and those that don't. Fu

3.4.7.4 Curve Signature

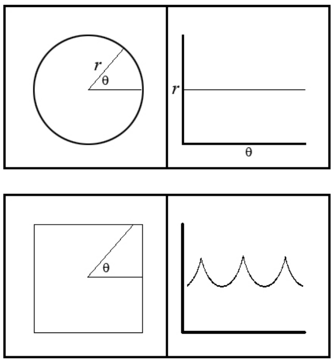


Figure 3.6: The top figure illustrates the signature of a circle which is simple a straight line with value r for all values of θ. The bottom is a circle which has a periodic signature with period $\frac{\pi}{2}$. The diagram follows that of Gonzalez and Woods [2001, p.648].

The term signature is used in different ways in the literature when referring to shape description. This study follows a variant on the definition in Gonzalez and Woods [2001, p.648] who define it as the distance between a pixel and the centroid as a function of the angle between them, as illustrated in figure 3.6. We modify Gonzales and Woods' definition in practice and

[16] i.e. with the variable ρ integrated or summed out of the expression.

define the signature as follows. Given the matrix of the skeletonised Image, $\mathbf{BW}(y, x)$, the signature may be defined as

$$\mathcal{S}(\theta) = \sum_{x=1}^{X} \sum_{y=1}^{Y} (\delta(\theta, \hat{\theta})) \sqrt{(x - O_x)^2 + (y - O_y)^2} \tag{3.5}$$

for discrete values of θ where O is the centroid of the skeletonised image, δ is the Kronecker delta function and $\hat{\theta} = (\tan^{-1} \frac{x-O_x}{y-O_y})$. This is the same as Gonzales and Woods' version except that our algorithm sums all 'on' pixels for every angle whereas Gonzales and Woods' definition only worked on closed boundaries.

The signature defined in this way is translation invariant, can be made scale invariant by taking its z-score, and a rotation of the image results in a translation of the signature.

3.4.7.5 Quadrant Statistics

This is a technique inspired by a method sometimes used in handwriting analysis of dividing the image of each letter into sectors and applying tests to each of these sectors and combining the results as a single feature vector. In quadrant statistics, the image is divided into four quadrants by a horizontal and a vertical line that intersect at the centroid. Then in each of the quadrants the following measures comprise the feature vector from that quadrant.

1. the position of the top pixel in the quadrant
2. the position of the bottom pixel in the quadrant
3. the angle between these two pixels
4. the ratio of the straight line distance between these pixels and the actual number of the pixels in the quadrant.

In the case of the top and bottom pixels in each quadrant, where there are more than one top or bottom pixels, disambiguation is achieved by selecting the top or bottom pixel furthest from the centroid. All four feature vectors are combined in a single vector such that elements 1 - 6 are from quadrant 1, 7-12 are from quadrant 2 etc, resulting in a 24 element feature vector.

3.4.7.6 Secondary Feature Extraction

Apart from the quadrant statistics all the feature vectors were resampled using Matlab's resample function so that they are all the same length for convenience. The choice of number of elements is somewhat arbitrary and 110 was chosen for all the primary feature vectors in this chapter (except for the quadrant statistics). Finally, PCA was used to reduce the dimensionality of the feature vectors. The number of components is chosen in two ways depending on the experiment and this is explained in the description of the respective experiments. Briefly, some use the virtual set as a validation set to select the optimal number of components while in other cases, PCA90/70 (explained in 2.3.2.1) is used to choose the number of components.

3.4.8 Results

The results suggest the following: Virtual samples can be used alone to train a machine to recognise form sets associated with a particular painter, real samples are slightly better than virtual samples (but not significantly so), using virtual samples to increase the size of the design set does improve the performance of all the classifiers, and majority vote substantially increases the performance of almost all the classifiers used in this experiment. A closer analysis follows:

3.4.8.1 Experiment 1

The results for the first two experiments are tabulated in table 3.5 and 3.6. The results suggest that when PCA90/70 is used, all three classifiers applied to any of the feature-spaces produce mediocre results when trained on the real set, with the exception of the vertical projection which is poor for all classifiers, and quadrant statistics which performs reasonably well for all classifiers. The good performance of the quadrant stats may be because it is the only one of these feature extractors designed specifically to deal with the problem at hand, whereas the others are general shape descriptors developed for different problem domains. This is, however, mere speculation and the more important result here is to establish a baseline which reveals mediocre to relatively good error rates on these test sets.

3.4.8.2 Experiment 2

The purpose of this experiment was to determine whether there is any real difference in performance between classifiers trained on a real design set and those trained on a virtual one and also to determine whether using the virtual set to select the optimal number of principal components yields an increase in performance over PCA90/70. Table 3.5 shows the point estimates and 95% upper bound for the error of 15 classification rules (i.e. LDA, QDA and 1-nn each trained on 5 different feature spaces) using PCA90/70, and table 3.6 shows the performance of the classification rules generated by the same classifiers but where the optimal number of components was chosen by using the virtual set for validation. When PCA90/70 is used there is only one classifier for which the upper bound on the real set error is lower than the point estimate for the error on the virtual set: when the nearest neighbour rule is applied to vertical projections. Table 3.6 shows that there is also only one such case when the virtual set is used to select the optimal number of components: when the vertical projection is used to train QDA.

The data are difficult to interpret in this raw form and to determine whether there is any significance to these differences, more sophisticated statistical tests are required. First, the McNemar test was used to determine whether there are any significant differences among the 15 classification rules between real and virtual sets. The χ^2 results are tabulated in table 3.7 and 3.8. Here, as well as for subsequent McNemar tests, a value of 3.815 or more represents significance at the $\alpha = 0.05$ level, and a reading of 0 simply means that the denominator in the McNemar comparison was < 10 and the result was discarded as being unreliable. Even without correction for multiple hypotheses, it appears that only one of these differences is significant - the nearest neighbour trained on horizontal projections when the optimal components are chosen using the virtual set for validation. However, since this is the only significant result among 15, the Bonferroni correction is appropriate, and using this correction (the modified χ^2 score should be above 8.62). Therefore, there is no conclusive evidence that any of the classifiers performs worse on real sets than on virtual sets.

While there is no conclusive evidence for any one classifier being improved significantly by using real data for training, it is possible that most of the classifiers are improved by a small amount, but that the sample size is too small to assess the significance of this improvement. In such a case, one may test for an overall improvement across all 15 classification rules. To do this, the Wilcoxon signed rank test was used. The results are that there is a borderline significant difference in performance at the $\alpha = 0.05$ level when using PCA90/70 ($p = 0.051$) and when using the virtual set for validation ($p = 0.063$). However, over all 30 classification rules, it appears there is a significant difference between virtual and real sets ($p = 0.007$). However, while there is a statistically significant difference, clearly in practical terms using virtual data appears to be almost as good as using real data.

In addition, it is worth determining whether using the virtual set as a validation set improves the performance of the classifier compared with PCA90/70. Comparing tables 3.5 and 3.6 suggests that the performance of LDA applied to the vertical projection is improved by using the virtual set to select the optimal number of components, regardless of whether the real set or the virtual set is used for training. To determine whether there is a statistically significant increase, both McNemar's and the Wilcoxon test were used. The results of McNemar's test (in table 3.9) reveal there is indeed a significant difference in performance for LDA trained on the vertical projections. There also appears to be a significant ($\alpha = 0.05$ corresponds with a χ^2 value of over 3.815) difference when QDA is trained on the horizontal projections. However, this last difference may simply be co-incidence, since there are 15 multiple hypotheses being tested simultaneously (using Bonferroni's correction, it would be rejected, but the first would not). To determine whether there is an overall difference across all 15 classification rules, the Wilcoxon signed rank test was again used for comparison. The results suggest that there is not a significant improvement either when using the real or the virtual design sets ($p > 0.05$ in both cases). Again, however, over all 30 classification rules there appears to be a statistically significant improvement for the validation set over PCA90/70, but in practical terms, apart from LDA trained on vertical projections, using PCA90/70 on the one hand and optimal components on the other amount to more or less the same performance.

3.4.8.3 Experiment 3

Experiment 3 is evaluated qualitatively rather than quantitatively by comparing the performance of each of the classifiers as more virtual samples are added. Figure 3.7 shows the graphs of the performances of LDA, QDA and 1-nn respectively on each of the five feature sets versus the number of virtual samples added to a design set composed of real samples, and figure 3.8 shows the mean performances of these 3 classifier types as a function of the number of virtual samples added. It is clear that there is a trend for all of the classifiers, not simply the complex ones, for an improvement as virtual samples are added to augment the design set. However, the improvement is minor, and as will be evident from experiment 4, this is dramatically overshadowed by the improvement due to using a majority vote ensemble.

3.4.8.4 Experiment 4

The final experiment was to determine whether combining the results using an ensemble would produce an improvement in the performance of the classifiers. The overall error rate from the ensemble is 0.07 (with a 95 % upper bound of 0.15). This is a significant improvement over any of the individual classifiers (from table 3.6). Table 3.10 uses McNemar's test of significance to compare the ensemble performance against those of the individual classifiers when trained on a combination of the real and virtual design sets (but evaluated only on the real samples using leave-one-out cross-validation). The table lists the scores (again, $\alpha = 0.05$ significance level is a χ^2 score of 3.815 for a single test and 8.62 when using the Bonferroni correction, and a 0 meaning the test was rejected) and there can be little doubt that the majority vote ensemble provides a very significance performance increase over almost all of the base classifiers (except where the data were insufficient for a conclusive result). In the subsequent chapters, the ensemble method will be the first line of attack.

3.5 Conclusion

Morelli's method has been the backbone of attribution studies during the 20th century, albeit in some or other modified form. The idea is that certain methods of rendering minor details are representative of particular artists and by examining these minor details one may eventually arrive at an attribution. This chapter demonstrates, using the incisions on the knees of male figures, that a computer may distinguish between different form sets. In particular, three form sets were defined, each of which was a particular method of rendering the incisions on a knee associated with a particular agent: in this case Exekias, Group E and the Princeton Painter. In addition, it was shown that such a system could be trained on drawings by an art historian in the manner of the particular painter, and furthermore, that when using a wide range of shape descriptors and a number of different classifiers, majority voting between these combinations yields very low error rates.

Interpretation of these results requires some discussion for a number of reasons. First, it is important to point out that attributing to form sets associated strongly with one artists is not the same as attribution to an agent associated with the artist. In the first instance, the agent is meant, in most cases, as a model of the artist's as a "black box" artistic machine, whereas a form set is one step removed from this. This means that interpretation of the results of any such test by a machine must be considered together with other results, or should be considered by an art historian. A second caveat is that this method is not meant, by any means, to be an objective method for identifying a painter, but rather a computer implementation of a method which is used by art historians. There is no assumption made that Morelli's method is valid, but rather that if it is valid, using shape descriptors to identify members of Morellian form-sets may be a useful method.

However, despite this caveat, the technique holds a lot of promise for a practical attribution machine that could be used in the field. The main reason for this is that it is relatively easy to develop a database of suitable training data since we have shown that reasonable results may be obtained on artificial samples. The implication of this is that experts on individual artists may construct design sets without requiring new photographs to be taken of the artist's *oeuvre* (let alone the 3D techniques suggested in 6.3). Of course, a database of higher quality (preferably 3D) images will only improve the system.

In addition, the error rate of 0.07 with the upper bound of 0.16 must be interpreted with two considerations taken into account. First, the test was conducted with a design set drawn by one art historian, and may reflect the skill or weakness of that individual (in this case an expert in the work of the Princeton Painter and his circle) at identifying the salient traits of the respective artist's form-sets. However, since the method proposed in this chapter is meant to be, to put it colloquially, an automation of the instinct of a skilled art historian rather than an objective method for identifying painters, artificial form sets by a skilled art historian are arguably more valuable, since they better encode the intuition of the art historian. For more objective analyses, methods should be sought that are more free of the bias of the art historian. The methods suggested in chapters 4 and 5 attempt, to some degree, to address this issue. The second consideration is that the design set was small as the aim was proof of principal rather than any attempt at an optimal solution. Therefore, it is quite possible that there will be improvement on the system is larger design sets are used. This may allow for even more complex classifiers to be used in the implementation of the system. What is clear is that, while there appears to be a small difference between the virtual and real design sets, the virtual samples are clearly sufficient for the purpose of classification and increasing the number of virtual samples appears to increase the performance of all the classifiers. Given that the virtual samples are relatively easy to produce, there are few constraints on the potential size of a design set.

3.6 Figures and Tables

		H-transform	**Curve Sig**	**H-Proj**	**V-Proj**	**Q-Stats**
LDA	Real UB	0.29	0.53	0.48	0.66	0.25
	Real	0.18	0.41	0.37	0.55	0.15
	Virt UB	0.34	0.54	0.44	0.58	0.3
	Virt	0.23	0.43	0.33	0.47	0.2
QDA	Real UB	0.41	0.44	0.41	0.5	0.34
	Real	0.3	0.33	0.3	0.38	0.23
	Virt UB	0.46	0.51	0.53	0.59	0.32
	Virt	0.35	0.4	0.42	0.48	0.22
1-nn	Real UB	0.39	0.46	0.32	0.41	0.32
	Real	0.28	0.35	0.22	0.3	0.22
	Virt UB	0.41	0.48	0.48	0.56	0.25
	Virt	0.3	0.37	0.37	0.45	0.15

Table 3.5: Error ε and 95% upper bound (UB) of LDA, QDA and 1-nn trained using different feature extractors both on real and virtual data, using PCA90/70 to select the number of principal components

		H-transform	**Curve Sig**	**H-Proj**	**V-Proj**	**Q-Stats**
LDA	Real UB	0.32	0.48	0.38	0.44	0.25
	Real	0.22	0.37	0.27	0.33	0.15
	Virt UB	0.36	0.5	0.41	0.39	0.3
	Virt	0.25	0.38	0.3	0.28	0.2
QDA	Real UB	0.34	0.34	0.36	0.51	0.34
	Real	0.23	0.23	0.25	0.4	0.23
	Virt UB	0.41	0.44	0.36	0.64	0.29
	Virt	0.3	0.33	0.25	0.53	0.18
1-nn	Real UB	0.38	0.42	0.25	0.44	0.29
	Real	0.27	0.32	0.15	0.33	0.18
	Virt UB	0.41	0.48	0.36	0.44	0.25
	Virt	0.3	0.37	0.25	0.33	0.15

Table 3.6: Error ε and 95% upper bound (UB) of LDA, QDA and 1-nn trained using different feature extractors both on virtual and real datasets using the optimal number of components established from validation using the virtual set

Classifier	**H-transform**	**Curve Sig**	**H-Proj**	**V-Proj**	**Q-Stats**
LDA	0.	0	0.04	0.6	0
QDA	0.2	0.4	1.7	0.9	0
1-nn	0	0	3	2.5	0.7

Table 3.7: McNemar χ^2 values for error rate comparison between Real and Virtual design sets using PCA90/70 for both.

Classifier	**H-transform**	**Curve Sig**	**H-Proj**	**V-Proj**	**Q-Stats**
LDA	0	0	0.5	0.4	0.06
QDA	0.6	0.2	0.07	2	0.6
1-nn	0.07	0.2	7.6	1.4	0.08

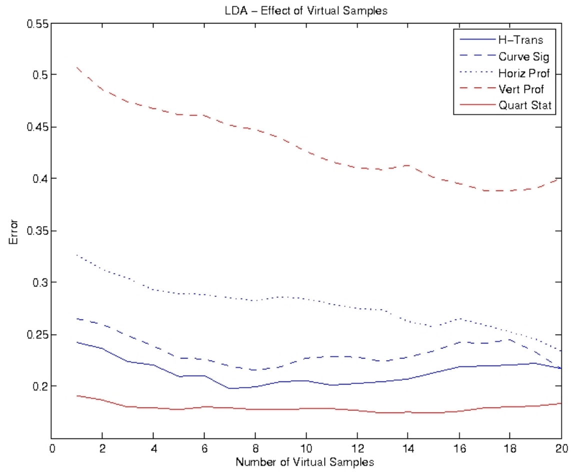
Table 3.8: McNemar χ^2 values for error rate comparison between Real and Virtual design sets, with components for both sets selected using optimal components

Classifier	**H-transform**	**Curve Sig**	**H-Proj**	**V-Proj**	**Q-Stats**
LDA	0.4	0.7	1.3	19	0
QDA	2.1	0	6.3	0.35	0.04
1-nn	0	0	0.6	0	0

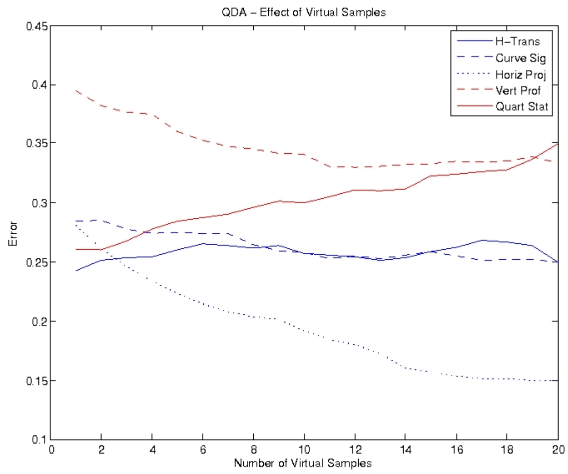
Table 3.9: McNemar χ^2 values for error rate comparison between PCA90/70 and optimal components, across all 30 classification rules (i.e all features and with both Virtual and Real design sets.

Classifier	H-trans	Curve Sig	V-Projection	H-Projection	Quart Statistics
LDA	0	17	5.8	16.4	0
QDA	12.5	7.7	0	11.3	0
1-nn	12	8.6	2	9.6	0

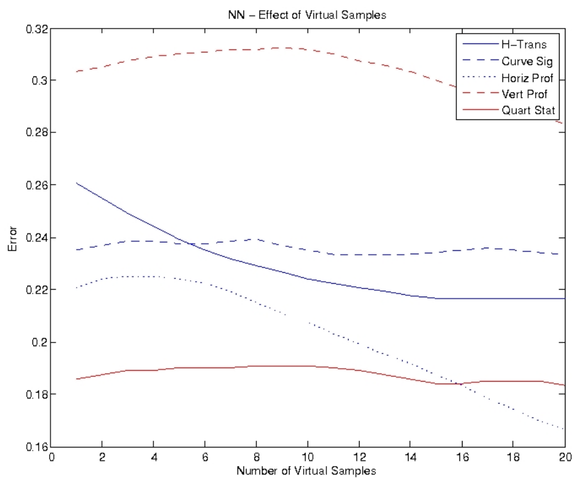
Table 3.10: McNemar χ^2 scores for majority vote versus all individual classifiers



(a) LDA

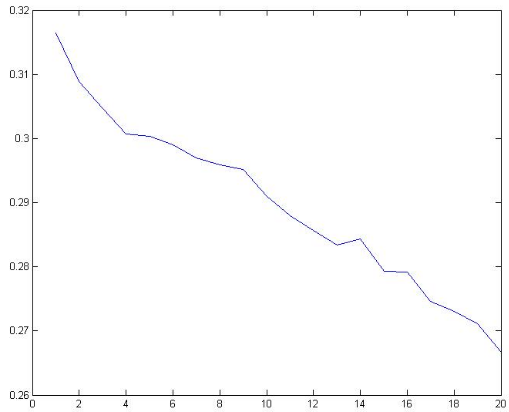


(b) QDA

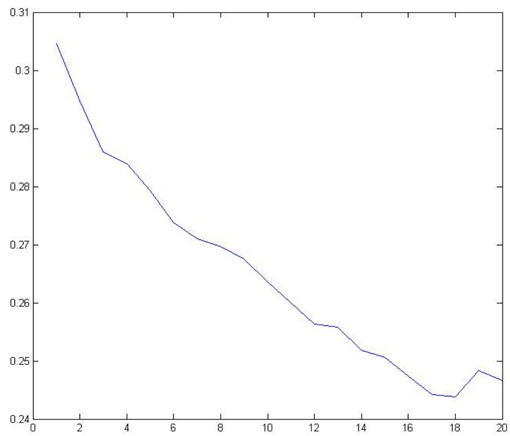


(c) 1-nn

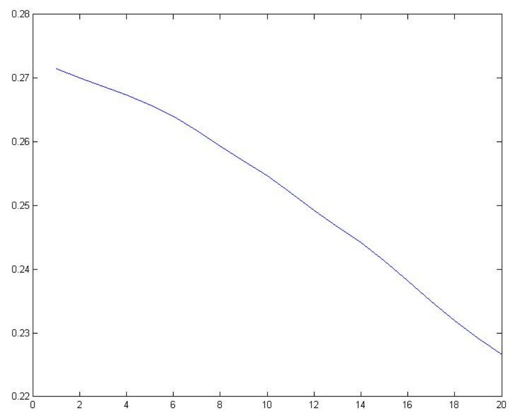
Figure 3.7: Plots of accuracy versus number of virtual samples added to the real design set .



(a) LDA



(b) QDA



(c) 1-nn

Figure 3.8: Plots of mean accuracy versus number of virtual samples added to the real design set.

Chapter 4
Cranial Proportions

4.1 Introduction

The previous chapter proposed using Morelli's method as the basis for a computer-assisted attribution system. However, this system still requires considerably judgement on the part of the human expert who is first required to identify what form-sets are associated with which agent. Morelli's technique is not the only approach to the study of attribution. In fact his scientism has come under attack for presenting only one facet of the attribution task, and concentrating on only a small aspect of artistic style. For example, critics have pointed out that certain less easily quantified properties, such as quality, proportion and balance, are ignored by the Morellian method. Among his successors, Berenson used subjective criteria as part of his methodology as did Friedländer [1960] who explained this subjectivism in terms of Gestalt psychology which claims that the expert would know the attribution from many subtle clues and simply use the empirical method as a post-hoc justification.

The terms proportion, balance and symmetry are frequently used in art historical discourse, but seldom in such a way that allows them to be easily understood in the formalisms required for pattern classification. In the first instance, quite often the discourses surround prescriptive quests for ideals, such as the divine ratio, to which some scholars believe good art should always adhere, rather than to find out what deviations from these ideals tells us about the art. In the case of attribution studies, aesthetic concerns are largely irrelevant with the prime interest being the discovery of what makes painters different rather than normative statements on what should unite them. Secondly, the terms are not used in a consistant manner. For example, even though the term proportion has traditionally been used to describe issues surrounding perfect quantities, it is often clear that the term proportion carries a number of connotations apart from the mathematical one. The following quote shows that, for Arnheim at least, proportion is not simply an issue of ratios, but says something about an innate sense of correctness that is dynamic and related to certain undefined forces.

> "The sense of proportion is inherent in the experience of perception, and - like all other perceptual properties-it is dynamic: rightness presents itself not as dead immobility but as the active equipose of concerted forces while wrongness is seen as a struggle to get away from an unsatisfactory state."[Arnheim, 1955, p. 44]

In this chapter the concern is not with the question of whether there is some universal sense of proportion and balance that is shared by all humans and which is a prerequisite for artistic perfection. Instead we are interested in whether there is a sense of proportion that is unique to each individual and which is manifest in their creative output. Furthermore, for this study, proportion is understood in its narrow sense of ratios between measures and specifically in this case, the measures of distance and angle are both considered.

In particular, a novel method is presented whereby the relative proportions, as defined above, of certain anatomical features are used as training data for classifiers to distinguish between the works of different painters. Doing so requires anatomical details that are invariant to much of the whims and conscious decisions of the painter and are instead rendered unconsciously through practised manoeuvres. It is shown that a classifier may be trained on the relative positions of cranial features of male heads to recognise the agent responsible with significantly better than chance accuracy. Furthermore, evidence is presented that the same features may also be used as a measure of stylistic difference between painters as there is a high correlation between the confusion matrices and traditional art historical notions of stylistic difference between the particular agents chosen for this study. In addition, the study also suggests that the method may be generalisable as a similarly high accuracy is achieved applying the technique to heads of print-makers in the Japanese Ukiyo-e tradition.

4.1.1 Motivation

A set of features that might reveal a painter's innate sense of proportion and balance should ideally possess certain properties that will facilitate their use given the limited amount of data. First, if we assume the painter's sense of balance reflects some "subconscious" aesthetic sensibility, then the conscious choices of the painter must be treated as noise. Thus the feature set should be as invariant against a painter's conscious decisions, such as emotional expression, as possible. Secondly, since no automatic image processing is possible, we seek a feature space that may be manually pre-processed or labelled relatively systematically and objectively. In other words the features should be such that manual processing them or labelling will not be likely to influence the final results. The heads of male figures that are not covered by helmets or headgear are features that meet these criteria and appear *a priori* well suited forms on which to define a feature space for computer aided attribution. The rational is articulated below.

It may appear counter-intuitive to a modern reader that the way in which human heads are rendered should not be subject to the painter's conscious decisions. This is because human emotion is communicated strongly by facial expressions and consequently the way in which these features are rendered in paintings of the western tradition (from the Renaissance onward) is often very important to the artist as a vehicle for delivering an emotional statement to the viewer. Because the emotion intended to be conveyed varies a great deal from painting to painting, there is considerable variation in the way the facial features are rendered within the corpus of a single painter in these traditions. Furthermore, the angle from which the heads are viewed varies considerably, particularly given the freedom of the human form allowed in many art movements since the Renaissance. A final factor influencing the amount of variation in the way facial features are rendered in western art is that in traditions that strive for realism, the facial features are also subject to differences between the human models used by the painter for each painting.

However, in strongly iconographical traditions such as Attic black-figure, many of these sources of variation are absent. In fact, for the most part, heads are free of emotional expression which is instead conveyed by the gestures. In addition, human heads in Attic black-figure are almost always rendered in profile, which ensures that the same facial features are, for the most part, visible on all heads that are not covered by a mantle or helmet. Finally, the Attic black-figure tradition is classicised and for the most part not realist. Instead, the heads conform very closely to a standard ideal or prototype, a typical example of which is illustrated in figure 4.1.[1]

By contrast, other gross anatomical features such as the torso and the limbs are presented in a variety of poses and consequently subject to too much variation for consideration as a stylistic feature. This is not particularly surprising since gestures often convey meaning and expression in this visual tradition, instead of faces which are impassive. Since the head is invariant to much of the conscious thought of the artist, it may be a good candidate for attribution. Furthermore, there are oblique references to the manner in which certain painters render their heads that suggest that Beazley also saw them as attributable criteria. For example, of the Swing painter Beazley commented "his men all look like Geese".

Only uncovered male heads have been used as the vase-painting examples in this study, although both female and male heads have been used in the case of heads from real and Japanese prints. There are a number of reasons for the omission of female figures in the Greek examples. First, females are rendered in a very different way from men in vase paintings, most notably in that their skin is white while that of the males is black. Incisions, which often appear white-ish, stand out strongly against black areas on a vase and therefore, for the most part, details on male heads are incised. This is not as simple with females because incisions on white do not stand out very well.

[1] Detail from the Reggio fragments by Exekias, *ABV* 145.2.

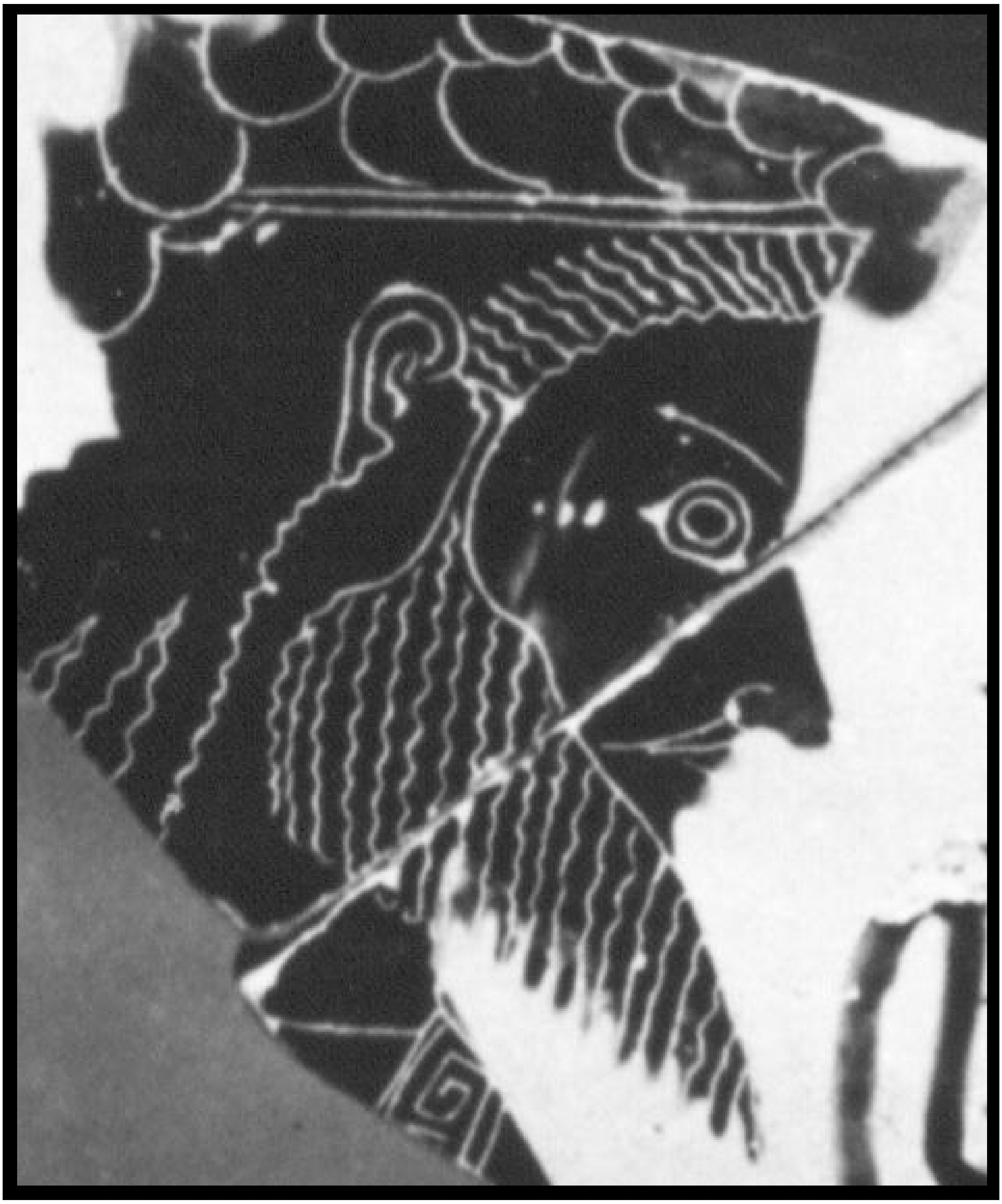


Figure 4.1: A typical Attic black figure male head. The figure is Dionysus, the god of ecstacy, wine and revelry. No indication of this nature is evident in his expressionless face, and instead it is conveyed by his iconography, particularly the wine-cup he carries.

Figure 4.2: Two images indicate the problem of analysing female heads when mostly black-and white photographs are used. The colour image reveals incisions nicely, whereas in the black-and white photograph these incisions are not visible. Notice that the male head on the left is clearly defined despite being photographed from a more oblique angle than the females.

Artist have used a number of techniques to get around this. Amasis, for example occassionally paints features. In general, female faces are rendered very differently in black-figure and consequently this study could not use both. The reason for choosing male faces over female faces is twofold. First, males are more common subjects in Attic vase paintings and therefore the central figure in the scene is almost always male. Most close-up photographs of vase paintings are of the central scene since bystanders are usually of little interest, and as a result, the best photographs are invariably of male subjects. Secondly, as has already been mentioned, incisions on white do not always show up very clearly, and in black and white photographs, it is difficult to make out all females cranial features (figure 4.2). These problems do not hold in the other media - the Japanese heads and the human heads in profile which are used as comparanda (explained in more detail in sections 4.2.5 and 4.3.1).

4.2 The Problem

Because they are invariant to emotional expression and severely restricted in the number of angles from which they are viewed, and given that other gross anatomical details are rendered with a range of different clothing and in a wide variety of different poses, the heads of figures on Attic black-figure vase-painters may provide a good basis for attribution.

This chapter has a three-fold aim. The first is to present a pattern recognition system that can potentially aid in the attribution of black-figure vase-paintings based on the relative positions and angles of cranial features of the male subjects. Secondly the system is used to determine which features are potentially the most useful for the purpose of classification. Third, for comparison, the technique is tested on works in another tradition in which the orientations are restricted: Japanese Ukiyo-e prints. Finally, two methods are used to determine whether the feature space may be used to define a measure of how closely painters worked together.

4.2.1 Method

2 sets of features are defined on the set of all uncovered heads of male figures in Attic black-figure vase paintings. The first is based on the relative positions of the major cranial features, described in section 4.3.3, and the second is based on the angles between these features. The data are features extracted from a set of uncovered male heads, 15 by the Princeton Painter, 12 by Exekias and 10 each by 3 other known painters,[2] and a control group of 10 real human heads photographed in profile. From these features, 12 classification rules are constructed by training LDA, QDA and 1-nn on the PCA90/70 measures of each of these features and to the z-scores of the PCA90/70 measures each of these features. This is illustrated schematically in figure 4.3. In other words, each of the three classifiers is trained on the principal components of the two original feature spaces, and also on the z-scores of the same feature spaces, giving a total of 12 separate classification rules. The final classification is made by a majority vote of the class predictions of all 12 classification rules. On this basis, four experiments are constructed. The first establishes a baseline measure of the error of the majority vote classifier and the performance of this classifier will serve as the proof of principal. The second aims to determine which of these cranial features are most diagnostic. The third determines whether these features have the potential to reveal how closely painters styles are to each other, both through an analysis of the

[2]The availability of decent photographs was a determining factor in the size of each set.

confusion between classes and also based on a distance metric defined on the respective feature spaces. Finally, the same procedure as experiment 1 is conducted on data from Japanese wood-prints to determine whether the approach may be general enough to apply to other artistic traditions.

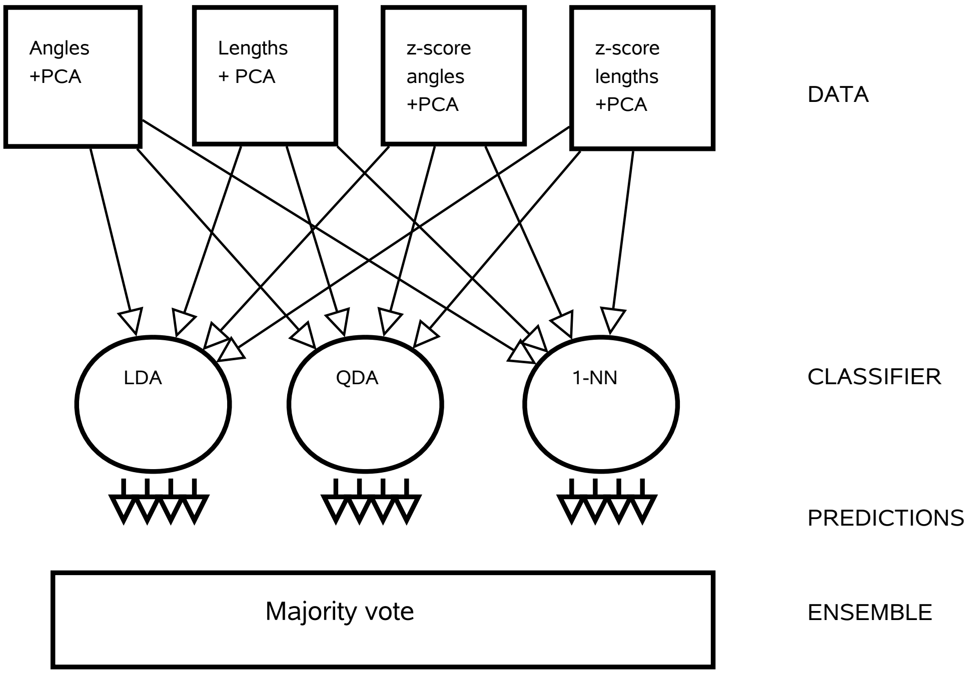


Figure 4.3: Schematic diagram of the classifier design. Both raw and standardised features from both feature spaces are used with 3 different classifiers to create 12 different sets of predictions.

4.2.2 Experiment One

The aim of the first experiment is simply to find a baseline performance measure for the system which will serve as a proof of principal. The basic principal is as described above - the 12 component classifiers each predict the class membership of each of the members in the training set using the leave-one out method (LOO). Thus for each object in the set, there will be 12 predictions of its class membership. An unweighted majority vote of these 12 prediction is taken for each object in the set, and the result is the final prediction of the classifier. The performance is reported both as the mode of the error estimate and the 95% lower bound.

4.2.3 Experiment Two

The second experiment attempts to determine which of these features are most responsible for the success of the classifiers and therefore which are most likely to be good candidates for use in subsequent research. To do this, experiment one is conducted multiple times each time with a different feature omitted. The relative performances are measured, and if any features are more diagnostic than the others, then the performance of the system would be expected to drop when these features are omitted from the feature space.

4.2.4 Experiment Three

The third experiment determines whether the method allows the computer to assess which painters are stylistically similar. The motivation is that, when painters work together, they pick up not only the broad artistic motifs of that workshop but also subtly influence each other's style. We know, for example, that some of the forms-sets used by the Princeton Painter are again used by the Swing Painter, such as the mouth being rendered as a single line, and the omission of weapons in armed combat scenes. More subtle elements of style are also absorbed by painters working close together to the extent that it is sometimes difficult to distinguish between paintings made by master and pupil. To this end, it is of some interest whether features used in these experiments are not only able to distinguish painters with a reasonable degree of success, but also whether these features reflect subtle similarities between painters working together.

Two between-class similarity measures are constructed, one based on an analysis of the classifier confusion, and one based on the distance between class means in the space formed by the 3 canonical variables[3] with the highest eigenvalues. These are compared with what is generally accepted as the stylistic relations among these painters. Table 4.1 shows the elements of the matrix P of the author's predictions of the similarity between different painters (1 is perfect similarity 0 is no similarity). This is admittedly the author's own subjective interpretation of established art historical opinion based primarily on Beazley's observations and also on those of specialists on the agents concerned. A rationale is provided in the discussion on each painter in 4.3.1. The table reflects similarities rather than differences so a distance matrix is defined: $H(i,j) = \frac{1}{P(i,j)}$. This experiment aims to determine whether the computer will concur with these differences. Thus H is compared with two different distance matrices defined on the feature space as follows:

[3]Canonical variables are explained in 2.4.3.1.

	Exekias	Group E	Princ P	Swing P	B 1686	Real H
Exekias	1	.6	.15	.1	.1	0.07
Group E		1	.2	.1	.15	0
Princeton Painter			1	.35	.3	0
Swing Painter				1	.25	0
Berlin 1686					1	0.05

Table 4.1: Predicted Similarities between different Agents (1 = perfect, 0 = disparate)

4.2.4.1 Confusion Analysis

For each of the component classifiers a confusion matrix is constructed. Briefly, a confusion matrix M is a matrix in which each element $M(i, j)$ represents the number of times an object that belongs to group i is classified to group j. In this experiment, the corresponding elements of the respective confusion matrices for each component classifier in the ensemble are summed up to produce a master confusion matrix. The confusion matrix is not symmetrical and therefore $M(i, j)$ is not a metric. Instead we define a measure $Q(i, j) = M(i, j) + M(j, i)$. Clearly $Q(i, j) = Q(j, i)$ so there will only be 15 unique such pairs if we exclude all entries $Q(i, i)$. However, the aim is to derive a distance measure, whereas high confusion actually implies similarity instead. Thus, a new matrix D is defined such that $D(i, j) = \frac{1}{Q(i,j)}$.

4.2.4.2 Differences in Group Means

For this experiment we define a feature space $\mathcal{L}$ as follows. If O is the set of all initial features (i.e. the normalised co-ordinates of the cranial features and the angles between them), and Z is the number of principal components that leads to lowest error (BLOO) in a LDA classification, then $\mathcal{L}$ is the space spanned by the 3 canonical variables (i.e. with highest eigenvalues) of the first Z principal components of O. The objects, represented in this space should tend to cluster in groups (a condition that is likely to be met if a linear classifier is able to distinguish between them). Here the implicit assumption is that stylistic difference should correspond to Euclidean distance between two objects in $\mathcal{L}$ and a measure of the difference between two painters is the standardised Euclidean distance between their respective class means.

4.2.4.3 Method of Comparison

The task of finding a suitable statistic to test the hypothesis that the distance measures defined above do indeed correspond with an art historical notion of stylistic difference is a difficult one. There are two important confounding issues. The first is that, even if all three distance measures do reliable reflect how closely artists worked together, their scales may not be commensurate and there is no real way of knowing how they should be scaled appropriately. A solution to this problem is to use a statistic based on ordinal values rather than cardinal ones. For example, one may use a rank ordering statistic such as the Wilcoxon signed rank test or Kendall's tau beta correlation test. This leads to the second confounding issue: that we are more certain of certain stylistic affinities than others, but a rank correlation test won't be sensitive to this. For example, while art historians would almost all agree that Group E and Exekias are more closely related stylistically than the Swing Painter and real human heads, the same certainty does not exist for the notion that the relationship between Group E and the Painter of Berlin 1686 is more close than that between the Princeton Painter and Exekias, even if this is suspected. Thus, while we may expect high correlations of the ranked pairs of which we are certain there may be low rank correlation between the pairs of which we are uncertain.

Thus, a qualitative evaluation may be best and this should serve as a pilot for future studies rather than a proof of principal in itself. In order to achieve this, a set of predictions is made that should be expected to hold if the hypothesis is correct that the two computer generated distance measures are roughly equivalent to what art historians might believe to be real stylistic difference. These are listed in order of importance. These are based on the author's own beliefs, but which are motivated by an interpretation of previous scholarship in 4.3.1:

1. Group E and Exekias should be more closely related than any other pair of agents.

2. The real heads should stand far apart from the other groups of painters

3. The Painter of Berlin 1686, Swing Painter and Princeton Painter should be closer to each other than they are to Group E and Exekias

4. Exekias should be closer to real heads than the other painters, followed by the painter of Berlin 1686

5. Group E, The Princeton Panter and the Swing Painter should be most different from the real heads

In addition to this discursive evaluation, a correlation is performed for the sake of completeness, but it should serve only as a guide. For each of the

3 methods (the art historical predictions, confusion analysis and difference in group means) a table of 15 pairwise ranks of the similarities between classes is established. For example, from table (4.1) the greatest similarity is that between Group E and Exekias, so this will be ranked 1, while the respective similarities between real heads on the one hand and Group E, Princeton Painter and Swing Painter on the other, are the greatest, so they will be ranked 13, 14 and 15. The Pearson correlation between between the ranks predicted by art historical criteria and each of the ranks calculated by the confusion analysis and the difference between class means is calculated. The significance of the correlation is determined by a Monte Carlo algorithm described in section (4.4.2)

4.2.5 Experiment Four

Finally, experiment four attempts to determine whether the technique works on artistic traditions in which there is a similar limitation on the number of angles from which faces may be viewed. If so, this will not only serve as a sanity check but also add some weight to the notion that the artistic personalities recognised by Beazley were in fact real artisans. There are, however, a number of differences between Attic vases and Japanese prints that are of some relevance to this comparison. The first is that the Japanese tradition, although it restricted the angle from which the heads were viewed, these are usually $\frac{3}{4}$ views rather than profile. This means that a different initial feature set needs to be chosen. Thus, the experiment does not test the particular feature space, but rather the general principal of using cranial proportions. Secondly, unlike the Attic tradition, the Japanese allows emotion to be conveyed by the facial expressions of its subjects. Nevertheless, the range of emotions is small and the nature in which they are displayed is highly idealised. In much the same way as the Greek tradition, the Japanese does not pay close attention to realism in the rendering of facial features, but is instead highly idealised.

The experiment is conducted along the same lines as experiment one. The co-ordinates of the relevant features are recorded manually and these are processed according to the procedure outlined in 4.3.5. 3 classifiers, LDA, QDA and 1-nn are trained on the data from each of the 4 feature extractors (co-ordinates, angles and z-scores of both) and the 12 sets of predictions obtained through the LOO method are combined using a simple majority vote. The error estimate and 95% upper bound are calculated in the same manner as experiment one.

4.3 Data

The data for the vase-painters is composed of relative positions of cranial features from between 10 and 15 paintings each by 5 different painters as well as a reference set of 10 real human heads, resulting in a total of 60 items. The choice of agents is influenced by three factors. First, experiment 3 assesses whether the method can be used to determine a "stylistic distance" measure. Testing this requires that the set of agents should include both those which are perceived by art historians to be distant from, and those perceived to be near to, the Princeton Painter. Secondly, the choice has also been motivated by availability of good quality images of heads. The author has a personal collection of photographs, and there are a small number of monographs on individual agents with good photographic plates. There is also the Beazley archive and *CVA* online, although the better agents tend to be better represented in these repositories. Finally, since the Princeton Painter is the subject on which the dissertation's methods are tested, the agents chosen in this group should reflect some relationship with the Princeton Painter.

For the Japanese print-makers, the three most famous painters of their respective eras are chosen, and images chosen at random from the photographs available to the author.

4.3.1 The Agents: Greek Vase-Painters

With these criteria, in mind, the following agents have been made part of the set: The Princeton Painter, The Swing Painter, The Painter of Berlin 1686, Group E, and Exekias. Below is a brief discussion of the relevant issues concerning the agents.

4.3.1.1 The Princeton Painter

The Princeton Painter is the subject of the studies in this dissertation and has been discussed briefly in the introduction. In addition a catalogue of his vases may be found in the appendix (A.7). The Princeton Painter is listed in *ABV* under the chapter heading "Other Pot Painters" indicating the degree of insignificance which Beazley attached to this painter. Boardman referred to the "Other Pot Painters" as an area for the connoisseur presumably as a reference to the general lack of importance of this painter in the development of Attic black-figure. Recently there has been some renewed interest in the Princeton Painter and since Beazley's death there have been a number of new attributions by various scholars. 15 heads by the Princeton Painter are used.

4.3.1.2 The Swing Painter

The Swing Painter, like the Princeton Painter, is one of the agents that Beazley listed under the chapter heading "Other Pot Painters", a remark that presumably reflected Beazley's disregard for their importance in the development of Attic Black-figure. Beazley said of the Swing painter that he was probably a pupil of the Princeton Painter, such is the similarity of their styles.[4] Böhr [1982], who has written the only published monograph on the Swing Painter, denies the master-pupil relationship but acknowledges the close similarity of these agents to each other. From the perspective of the aims of experiment three, the Swing Painter should be considered the closest vase painter to the Princeton Painter. He is also an interesting subject for this particular study because the faces of his men are distinctive, something picked up by Boardman [1974, p.63][5] among others. 10 heads by the Swing Painter are used.

4.3.1.3 The Painter of Berlin 1686

Another of the "Other Pot Painters" is the Painter of Berlin 1686. His style is also very similar to that of the Princeton Painter but somewhat more archaic. His relationship with both the Princeton Painter and the Swing Painter has been acknowledged by Beazley and Maxmin [1979, 1986]. In particular, Maxmin sees the connection between these two agents as something that grows towards the end of the painter's career when he is influenced by the fashionable economy with which they render anatomical details. However, the Painter of Berlin 1686 is less close to the Princeton Painter than the Swinger is, and he is less close to the Swinger than he is to the Princeton Painter. 10 heads by the Painter of Berlin 1686 are used.

4.3.1.4 Exekias

Exekias was introduced briefly in the previous chapter. He is widely considered as the best of the Attic black-figure agents and has numerous publications dedicated to his *oeuvre*.[6] His style is refined, neat, detailed and, as much as the severe tradition would allow, realistic. Of all the black figure artists, Exekias' faces come the closest to having expression, something apparent in the scene of Ajax committing suicide in which the protagonist's

[4] *ABV* 304.

[5] "He is ... not a conscious comedian but his placid figures with their big heads, [and] fashionable tiny noses..."

[6] Monographs include Technau [1936], Mackay [1981] and Mommsen [1997]

brow is wrinkled as though contemplating his fate (figure 4.4(b)).[7] Exekias' realism is also apparent in the drawing of the black African phenotype on the attendant of Memnon in London B209[8] (figure 4.4(a)) in which he not only attempts to illustrate the platyrrhiny with a wrinkled and upturned nose, but in which the figure's hair is rendered in relief.[9] For this reason of all the artists in the group, we should expect the proportions of the heads of Exekias to be closest to those of real human heads. 12 heads by Exekias are used.

4.3.1.5 Group E

Group E, as has already been stated in 3.4.6.2, is a large group that is stylistically quite compact and closely related to Exekias. The technical mastery of this group of agents falls considerably short of that of Exekias, but slightly surpasses the Swing Painter and much of the work of the Princeton Painter which is often rushed. Traditionally, Group E and the Painter of Berlin 1686 are thought to be slightly earlier than the other three agents. The distance between Group E and Exekias is closer, by art historical reckoning, than between any other pairs of agents in this study. In addition, from the evidence of their choice of subject and the compositional templates there may also be some connection between the Princeton Painter and Group E - something I have argued in A.4.4. Nevertheless, this connection has not been noticed by other scholars and is unlikely to be reflected closely in their respective styles. 10 heads by Group E are used.

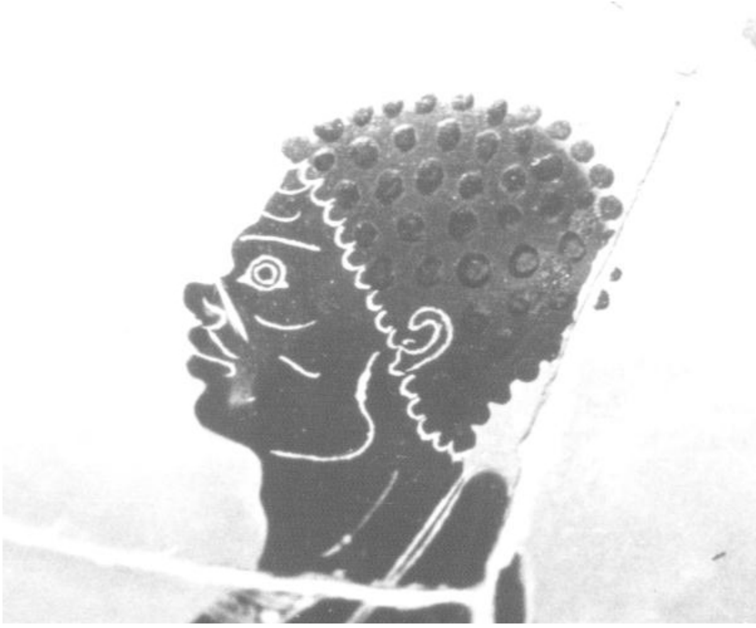
4.3.1.6 Real Human Heads

In addition to the heads of male figures in vase-paintings, a set of 10 real human heads in profile has been used as a ground-truth measure, and also to determine whether any of the vase painters is more realistic than any other. The heads are both male and female rather than simply male for two reasons. First, the proportions of human heads are so similar, that the difference between male and female heads is insignificant compared with that of the vase-painters. For this reason, a mixed group was believed to have more variation and provide a better set for comparison. As a ground truth comparison we should expect that the relationship between the real heads and any of the other agents will be considerably smaller than the relationships

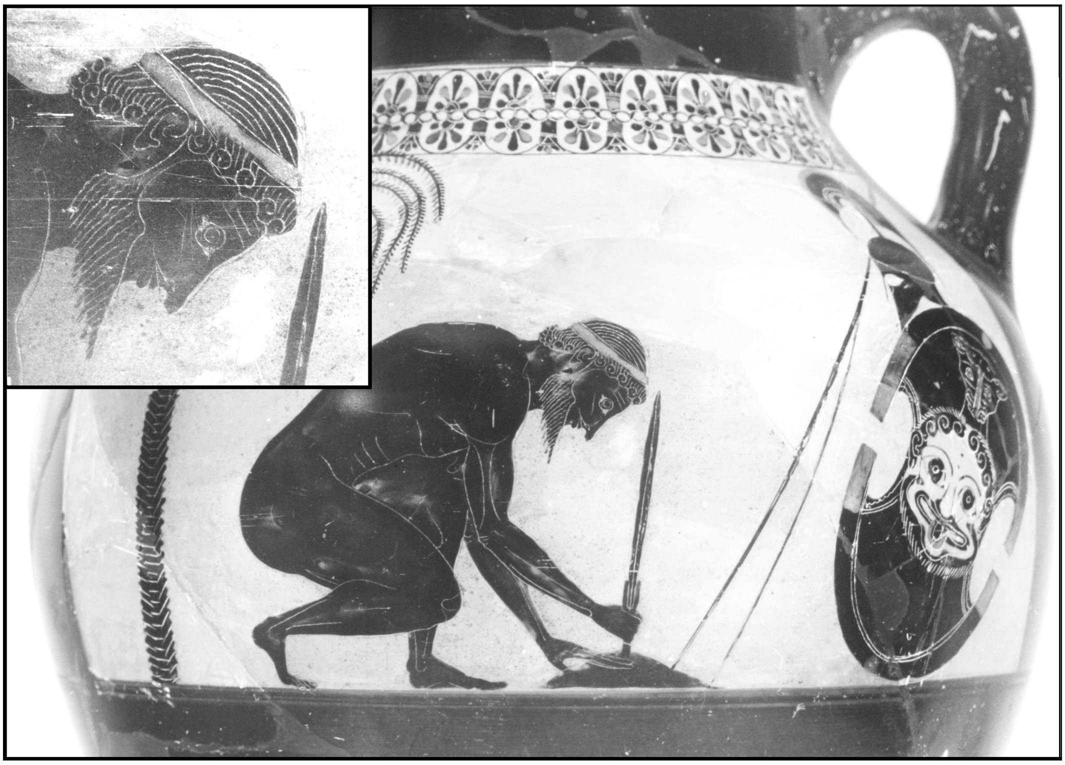
[7]Boulogne 558 *ABV* 145.18 *Para* 60 *Add*[2] 40.

[8]*ABV* 144.8, 686 *Para* 60 *Add*[2] 39.

[9]Noticed by Snowden in *LIMC* s.v *Aithopes* as proof that Greeks had close contact with black-Africa by the late 6th century BCE).



(a)



(b)

Figure 4.4: Two images by Exekias that illustrate his attention to detail regarding facial characteristics.

between any other two agents. Furthermore, we should expect that Exekias, who appears by art historical considerations to strive for naturalism, to be closer to the real heads than any other agent is.

4.3.2 The Agents: Japanese Print Artists

In addition to the Attic heads, the same technique is tested on heads from Japanese Ukiyo-e ("floating world") woodblock prints from the Edo period (1603-1867). The most popular subject of this tradition was actors from the Kabuki theatre, although a number of other subjects are well-represented, including landscapes, city-scapes and beautiful women (*bijinga*). Because of the subjects, there are many portraits and the heads of individuals are easy to come by. Unlike the Attic artists however, there is some expression in the paintings of the Japanese prints. The reason for their choice as a comparandum however is that in many cases, the expressions are themselves stylised rather than naturalistic and furthermore, that the viewing angle is usually $\frac{3}{4}$ view allowing comparisons to be drawn between paintings. The technique of Ukiyo-e involved painting a master drawing which was stencilled into a block used to make master prints. The different colours were each then stencilled onto different blocks and these blocks were used to print the different colours on the final works.

4.3.2.1 Utamaro (1754-1806)

Towards the end of the 19th century, the specialisation in human subjects began to wain in favour of landscape and other subjects. Utamaro is considered to be the most important Japanese print artist of the final phase of human subject specialisation within the Ukiyo-e tradition. While his earliest interests were in the theatre and actors, he later increased the range of subjects to include among other things, natural works. Utamaro is particularly well known for his studies of the female form and for his use of composition.

4.3.2.2 Toyakuni (1769-1825)

Toyakuni is one of the earliest print-makers of the Utagawa school, and a successor of the school's founder, Toyoharu. The school had introduced western techniques into Japanese art, particularly deep perspective, something best realised in architectural and landscape prints, and many of its adherents exploited this by creating city-scapes, rooms with exaggerated depth, and fantastical landscapes. However, Toyakuni eschewed the landscape for a return to the traditional Ukiyo-e subject of the actors and in particular Kabuki theatre. Unlike many of his contemporaries, Toyakuni did not paint portraits of actors in stereotyped poses, but copied actual scenes from the theatre.

4.3.2.3 Kunisada (1786-1865)

Kunisada was also a member of the Utagawa school and was the school's most prolific artist. He is also the most popular and commercially successful 19th century Japanese artist. Although his subjects were the traditional Kabuki actors, he painted on a wide range of themes including beautiful women and Sumo wrestlers. His style was particularly noted for his innovation at a time when the art was in decline.

4.3.3 Initial Feature Set

The initial feature is composed of the co-ordinates of various cranial features of the heads of painted figures. In the case of the Attic vases, these are all male heads in profile, whereas in the case of the Japanese prints, these are both male and female heads in $\frac{3}{4}$ view. Classification is not made on the basis of the initial feature set, but rather on features derived from this set (described in 4.3.5).

The initial feature set was chosen based on the following criteria. Each cranial feature had to be

1. representable by a single x-y co-ordinate
2. rendered on all heads in the set of examples
3. not subject to considerable change between paintings by the same painter

Point 1 rules out features such as the length of the hair or the curvature of the forehead. Point two rules out moustaches and filets (headbands). And point three rules out hairstyles (which change frequently). By these criteria, the following facial features were chosen for the Attic figures (illustrated in figure 4.3.3.2)

4.3.3.1 Features: Attic vases

1. the middle of the ear
2. the middle of the eye
3. the left corner of the eye

4. the right corner of the eye
5. the juncture of hair and forehead (i.e. the crown)
6. the bridge of the nose
7. the tip of the nose
8. the mouth
9. the chin/tip of the beard
10. the point of intersection of the back of the head and a line through points 1 and 8
11. the point of intersection of the face and a line through points 1 and 8

and for Japanese artists the following were used (illustrated in figure 4.3.3.2):

4.3.3.2 Features: Japanese Prints

1. the middle of the ear
2. the left and right corners of both eyes
3. the juncture of hair and forehead (at the horizontal mid-point)
4. the bridge of the nose
5. the tip of the nose
6. the middle of the mouth
7. the chin

4.3.4 Capturing Data

The data is captured from photographs from the author's own collection of Group E, Exekias, The Princeton Painter, The Painter of Berlin 1686 and the Swing painter. The head of the relevant figure was first cropped from the rest of the image. No scaling was necessary since the derived features are scale invariant. However, if the image was greater than 450 pixels in any direction, it was scaled down so that the maximum size was 450 - in order for the image to fit on the space provided by the software. In addition a set

Figure 4.5: The initial features: Greek Vases.

Figure 4.6: The initial features: Japanese Prints.

of heads taken from human figures in profile was used as a reference set. A proprietary C++ executable was written which displays heads by different agents and prompts the user to click on a specific feature. All heads were made to face to right by horizontal flipping if required, and in one case the head was rotated by 90 degrees to make it face right. The co-ordinates are recorded in a text file with the x and y co-ordinates recorded and separated by a line break, and each head separated by a ';'. The text file was imported into MATLAB where processing was conducted.

4.3.5 Derived Features

The raw co-ordinates were standardised as follows, so that meaningful comparisons between paintings could be made. First, the data were translated so that the origin of the data for each head was the co-ordinate for feature 1. Secondly, the data were scaled so that the respective distances between feature one and feature two were 1. Finally, the data were rotated so that the angle between feature one and feature two was 0. The space spanned by this standardised data comprised feature space one. The second set of derived features comprised of the angles between each of the points in feature space one. For each of the objects to be classified, four feature vectors were constructed. The first was the co-ordinates of the objects in feature space one, the second the co-ordinates in feature space two, and three and four were z-scores of feature space one and two respectively. The classifiers were trained on the PCA90/70 scores of these feature vectors. All the derived feature spaces are scale, translation and rotation invariant.

4.4 Results

4.4.1 Experiment One

The baseline result of experiment one suggests that the classifier is capable of distinguishing between the heads of various agents with above chance accuracy. On 60 heads, the classifier achieved an error of .36 with a 95% upper bound of .46 compared with a 0.8333 expected by guessing. On the 50 heads by the vase painters only, this drops to .41 with a 95% upper bound of .52. This is still very much better than chance for a 5 way classification (chance = 0.8) but not very good. Clearly the technique does work, although as it stands it is not a good classifier. One of the reasons for the low score is that the score on Group E is very low. This may reflect that Group E is not considered to be the output of a single artist, but of many artists working closely together, of which Exekias was probably a member (although his work is considered separately). Removing Group E from the set results in an error of 0.34 with a 95% upper bound of 0.47.

4.4.2 Experiment Two

A comparison between the ranks of the distances between agents based on the confusion matrix, the class means and the art historical predictions are shown in (table 4.2). Here a lower rank implies a closer relationship. On the main relationships they do agree. First, the closest relationship (rank 1) by all counts is that between Group E and Exekias which is probably the relationship that would find most agreement between art historians. Secondly, the real heads are very different from those of any of the vase-painters, but as predicted, Exekias fares considerably better than the rest with the exception of the Painter of Berlin 1686 who gets a good showing from both the class means and the confusion based distance measures. In addition, the two computer techniques saw a very close relationship between the Princeton Painter and the Swing Painter, and the Painter of Berlin 1686. However, they both considered the Princeton Painter to be closer to Group E than to the Swing Painter - quite counter to art historical intuition. Where the computer predictions differed most from those of the art historian is in the relationship between the Swinger and the Painter of Berlin 1686. Here Maxmin has noted a relatively close working relationship, albeit towards the end of the painter's career, the computer predictions would seem to disagree. On many other relationships the predictions of all three methods disagree. However, for the most part, these were areas in which art historians do not have any fixed opinions.

The dendrograms in figures 4.7(a), 4.8(a) and 4.9(a) visualise these relationships in an intuitively appealing way. The dendrograms are constructed using hierarchical agglomerative clustering with single linkage as the distance measure. The method is as follows: at level 0, all elements form distinct groups. Then groups are successively merged when the distance between their nearest neighbours is smaller than the distance between the nearest neighbour to any of the other groups. The y axis on the dendrogram represents the distance at which the groups were merged. The method was implemented using the MATLAB linkage function. Group E and Exekias are labelled as a distinct group on their own by both the computer predictions and the art historical predictions. In addition the real heads are more different from the rest of the agents than any of the other agents are from each other. However, while they all agree that the Swing Painter, the Painter of Berlin 1686 and the Princeton Painter are very different from Group E and

	Confusion	Class Means	Predicted
Princeton - Exekias	6	8	6
Princeton -Real	15	14	13
Princeton - B1686	4	3	3
Princeton - Group E	2	4	7
Princeton - Swinger	5	5	2
Exekias - Real	10	12	12
Exekias - B1686	11	6	8
Exekias - Group E	1	1	1
Exekias - Swinger	3	7	9
Real - B1686	7	11	10
Real - Group E	14	13	14
Real - Swinger	13	15	15
B1686 - Group E	9	2	5
B1686 - Swinger	12	10	4
Group E - Swinger	8	9	11

Table 4.2: Ranked distances between different agents using different distance measures

Exekias, they do not agree on how close these three agents are to each other.

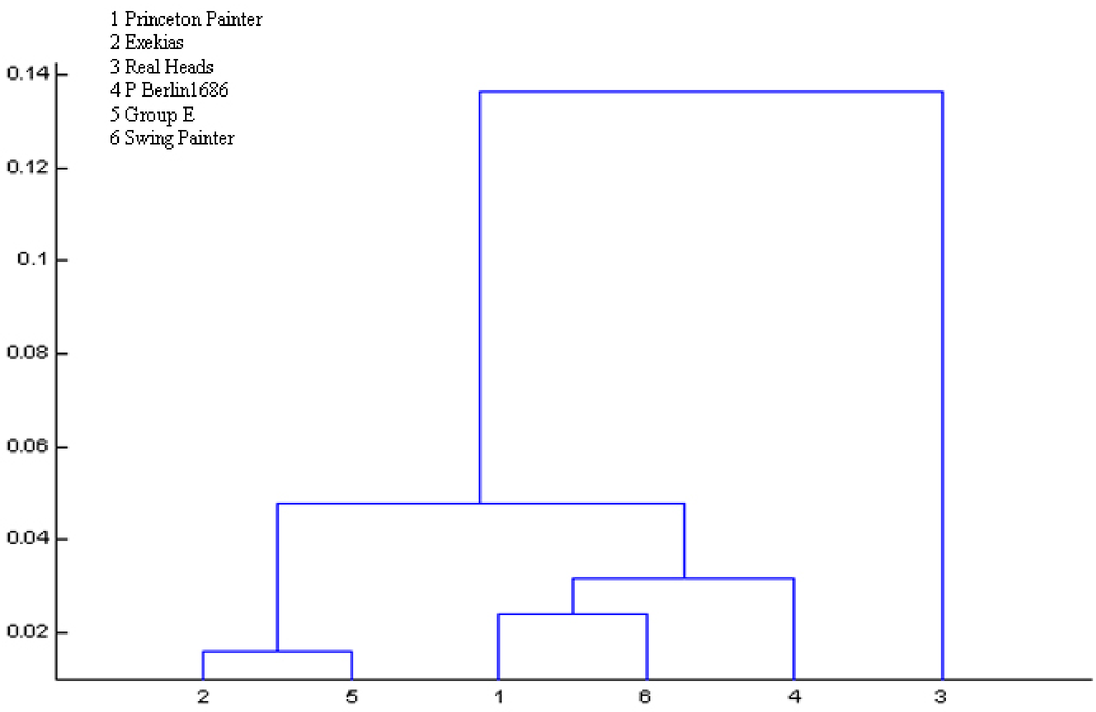


Figure 4.7: Dendrogram of class distances for heads by various agents according to art historical criteria.

The dendrograms are merely intuitively informative visualisation tools and do not allow us to asses whether the similarities between the sets of predicted ranks is significant. No suitable off-the-shelf statistic exists that can capture how closely the computer predictions match those of the art historian. A novel approach was used that employs a Monte-Carlo simulation to get a distribution of Pearson correlations between random vectors of predicted ranks and uses this distribution to assess the similarity between the art historical predictions and the two computer methods. First, for each of the following, the art historical predictions and the two computer predictions, a 15 element vector is constructed as described in 4.2.4.3. The elements of each vector are represented by the columns in 4.2 and are simply the ranked distances between the specific pairs of agents. The pearson correlation co-

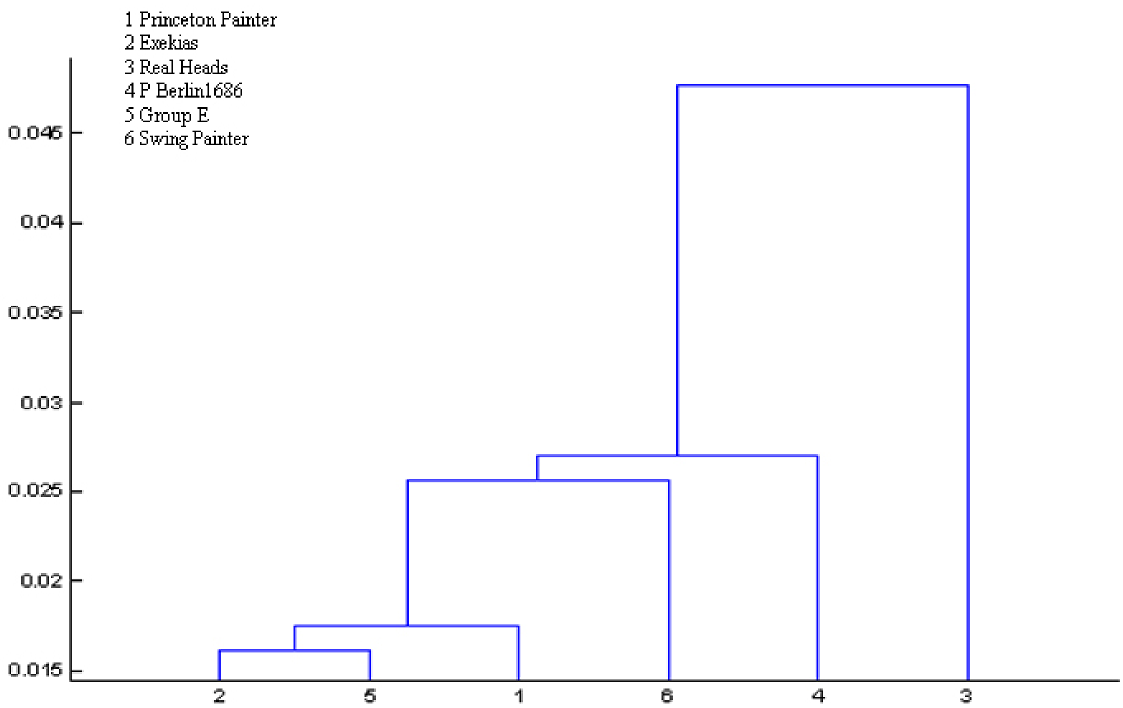


Figure 4.8: Dendrogram of between class distances for heads by a distance measure defined on the confusion matrix.

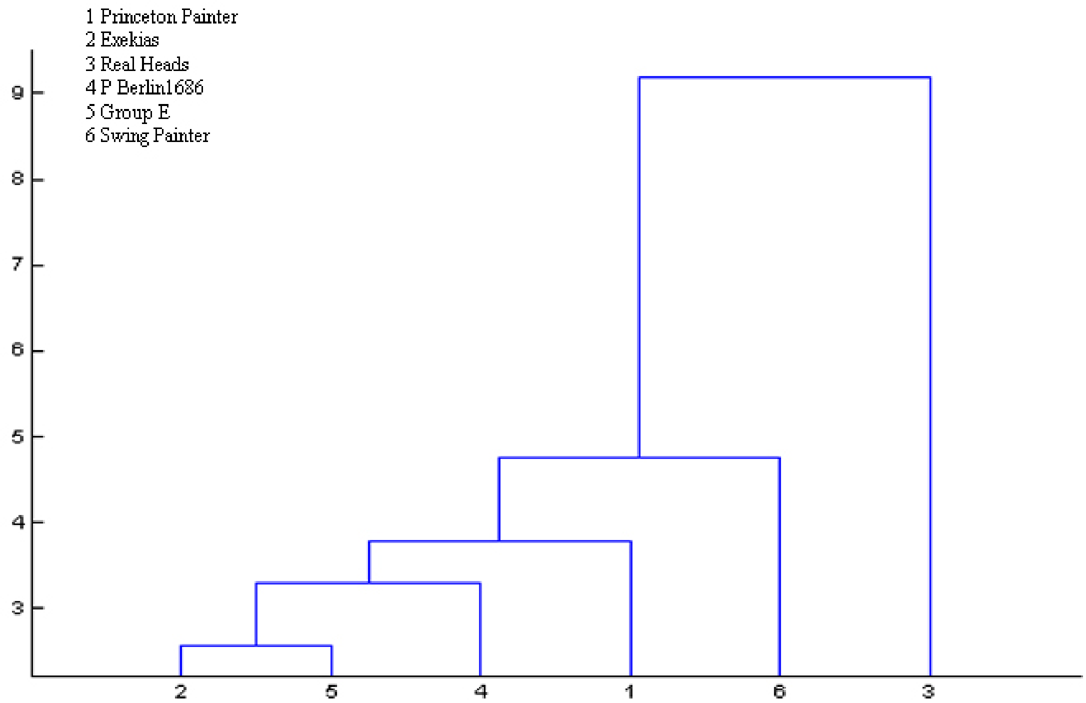


Figure 4.9: Dendrogram of between class distances for heads by various agents based on the Euclidean distance between class means in $\mathcal{L}$

efficients between the art historical predictions and each of the computer predictions is calculated. This co-efficient is a measure of how similar the art historical predictions are to the computer-based predictions. To convert this to a traditional p value showing evidence against the null hypothesis we construct a distribution of correlations co-efficients that would be expected under the hypothesis that the correlation between any two such rank vectors was due only to chance. To achieve this 2 vectors are constructed of 15 elements each, each element being a random number sampled evenly from the natural numbers between 1 and 15 without replacement, and the Pearson correlation co-efficient between these vectors is calculated. This process is repeated 10000 times to obtain a distribution of Pearson correlation values between two random vectors. The p value is simply the proportion of the 10000 Monte-Carlo values that are higher than the correlation between the two sets of ranked distances. The resulting p-values are 0.043 for the human predictions and class means on the one hand, and 0.001 for human predictions and confusion based distance on the other.

There are many confounding issues in the above analysis, and there are clearly many areas in which the computer calculated difference between two agents is not the same as that of the art historian. However, on the most crucial issues there is definite consensus: Exekias and Group E, The Princeton Painter and Berlin 1686, the Princeton Painter and the Swinger, and the real heads versus the rest. Furthermore, even though the statistical measures are imperfect, correlation statistics, suitably calibrated for this specific task, reveal that the correlations between the computer predictions and the art historical predictions are very strong and certainly not consistant with the hypothesis that they are due to chance.

4.4.3 Experiment Three

Figure 4.10 shows the standardised scores of the accuracies of an ensemble of six classification rules (LDA, QDA and 1-nn trained on each on z-scores and raw scores) when particular features are deleted from the feature-space. Correction for multiple hypotheses using Bonferroni's method,[10] none of the features appear to have any more diagnostic power than any other on their

[10] While the general use of the Bonferroni correction is often questioned (for example by Perneger [1998]), this is usually because it is so strict at controlling type I error that it can allow a large increase in type II errors, and it is argued that the only circumstances under which it is generally applicable is when the overall null hypothesis is of interest, the individual hypotheses are of considerable interest in themselves or (the Bayesian perspective) there is a prior belief that all null-hypotheses are true [Westfall et al., 1997]. The last two of these are true in this case.

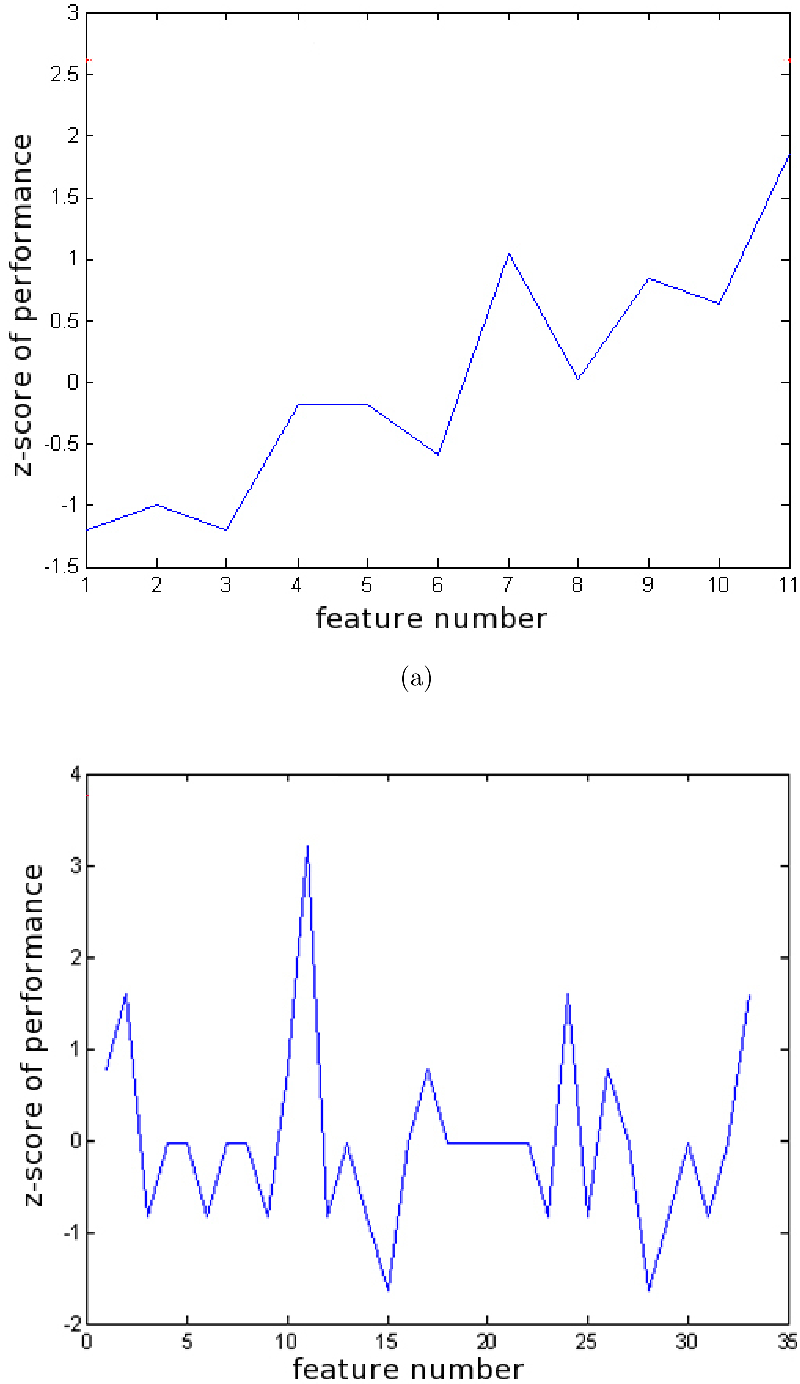
own, and common sense would dictate that they all be used in concert for optimal performance, unless a larger scale study reveals otherwise. Figure 4.10 does not label the features as it would reveal nothing since there is no significant performance difference between them.

4.4.4 Experiment Four

Experiment 4 reveals that the methodology of experiment 1 may be used to distinguish between faces of print figures from the Japanese Ukiyo-e tradition where the true identities of the artists were not in doubt. Here the classifier error was 0.24 with a 95% upper bound of .39 suggesting that the technique works on this tradition as well. This may be seen as a slight vindication of the notion that the agents identified by Beazley were indeed real agents. Of course, it is quite possible that Beazley's attributions were based, in part, on these very proportional relationships and therefore, this experiment must not be considered proof of Beazley's method but merely as supporting evidence, or at best, proof that it was not arbitrary.

4.5 Conclusion

This chapter is predicated on a belief that the relative proportions and angles between physical facial features as they are rendered on Attic black-figure vases are diagnostic criteria for artistic attribution. The motivation for this is that faces are stylised and expressionless and thus differences between agents will not be conscious and thus they may reflect ideosyncrasies of the individual artists' styles. To this end four experiments were conducted. The first was to test this hypothesis by training a recognition system on the positions of certain facial features that were considered most invariant to the whims of the painters. The error rate of the recognition system was estimated using LOO and the 95% lower bound taken as evidence for the hypothesis. Although the error was high, the classifier performed considerably better than chance, correctly assigning the heads of vase-painting agents 59% ($\varepsilon = .41$) of the time and 66% ($\varepsilon = .64$) when excluding Group E (which is not normally considered to be the work of a single painter). The second experiment found that on the most important issues, the distance between agents as measured by two computer based distance measures corresponded quite well with the distances predicted by art historical considerations. This suggests that this particular method may have further uses than attribution, such as verifying and establishing relationships among pot-painters, although further research is required. The third experiment revealed that none of the features could



(a)

(b)

Figure 4.10: Relative performance of classifiers (using standardised accuracy) with specific points deleted. (a) represents the positions of the cranial features. (b) represents the lines between these features. No features appear to be significantly more diagnostic than any others.

firmly be said to carry more diagnostic power than any other. Finally, as a sanity check, experiment one was conducted on heads of Japanese prints where the authorship was known beforehand. The system was quite successful on these heads lending some credibility to (but not proving) the notion that Beazley's agents may have been distinct artistic personalities.

CHAPTER 5

Vase Construction

The two previous chapters both introduced novel approaches to computer-aided attribution of art. The first was based on the method of Morelli which has dominated the discipline of vase-painting analysis. Because it was proposed as an empirical alternative to the traditionally subjective methods of Morelli's day, it naturally invites a computer-aided methodology. The second chapter used the proportions of anatomical details, specifically the relative positions of cranial features, as the basis for a computer based attribution system in the hope that these less intuitively quantifiable measures may not only be able to provide a basis for a classifier that can attribute art, but also might reveal insights into the attribution process. The final method proposed in this dissertation examines the shape of the vases to determine whether this may reveal the identity of the constructor of the vessel, both as an intermediate step in determining the painter, and because the identity of the potter is interesting in its own right.

5.1 Introduction

The production of Attic pots was the work of two different craftsmen (or artists) - the potter and the painter. The former was responsible for constructing the vase and the second only with decorating the surface. Perhaps owing to the art-historical roots of much of the modern discipline of vase-painting attribution studies, the construction of the artefacts themselves has not been studied in as much depth as the paintings on the surface. However, the manner in which the vases have been constructed offers important information for both attribution and also the dating of the artefacts. Two important factors underlie this. The first is that there exists a relationship between painter and potter that may often allow inferences about the identity of the one to be used to aid in the attribution of the other. For example, if painter A almost always paints vases potted by potter B, then finding that a pot has been painted in the style of painter A, with all else being equal, strongly suggests an attribution to potter B. The second is that, in general, vases (especially amphorae) get progressively slenderer over the course of the sixth century BCE. Thus this criterion may be used as part of the process of dating the artefacts.

It is difficult to determine with any certainty the exact nature of the relationship between potters and painters in Athens. In the first instance, there exists no extant ancient treatise that explains the production of, and trade in, vases. Instead, this information has to be inferred from the archaeological record. The primary evidence in this regard is the small number of signatures left by painters and potters, which together with careful analyses of the shapes of the pots, may be used to piece together the relationship between painter and potter.[1] The most common signatures that appear on vases are "[X] *egraphsein*" and "[Y] *epoiesen*" ("[X] painted (me)" and "[Y] made (me)"), or variants on these formulae. While *egraphsein* (painted) is not difficult to understand, *epoiesen* (made) is less obvious. While "made" would seem to suggest the constructor of the artifact - in other words, the potter - the term is used on different media such as sculpture and wall-paintings to indicate the name of the artists. It is conceivable then that a painter could sign with "*epoiesen*" instead of "*egraphsen*" and there is evidence that this is sometimes the case. In the first instance there are dual *epoiesen* signatures of the form "X and Y made (*epoiesen*)me", and since collaboration in the potting is exceedingly unlikely, at least one of X and Y must be the painter of the vase. Furthermore, many of the vases of Exekias are signed *Exekias epoiesen* and are clearly painted by him. It has been argued that *epoiesen* denotes the owner of the workshop, although there are cases where the term is used to describe a variety of objects that must have come from the same workshop, yet have different *epoiesen* signatures. Yet another alternative is that both potter and painter have the right to call themselves artists and thus both may sign using *epoiesen.* In most cases, however, the potter seems clearly to have been the person who signed *epoiesen.* Finally, there are also important cases is which potter and painter may be the same person, as is probably the case with Exekias (section 5.3.1.1) and may well be the case with Amasis (5.3.1.3)

5.1.1 Vase Shape and Attribution

Bloesch [1940] was the first to publish a methodology for attributing pots to their respective constructors by analysing their shape. His study was limited to cups but the scale was still very large - he attributed 900 vessels to the respective constructors. For Bloesch, the attribution of the vessel to the constructor was an end in itself, whereas this dissertation is mainly concerned with painting style. However, the constructor attribution is of considerable interest for painting attribution. If painter and potter are the same individual or if a single painter paints exclusively the work of a single potter, then an analysis of the shape of the vase may be used both for attribution or as a clue for the internal chronology of a painter's corpus. The first point is relatively straightforward, but the second requires further explanation. It has long been recognised that over the course of the sixth century BCE,

[1] As noted in Ch1 fn.23 only 40 artists sign their works and less than 1% of the extant vases are signed.

vases got progressively slenderer. While it is uncertain who was first to point this out, it is the basis of at least two studies of the internal chronology of Exekias[Technau, 1936, Mackay, 1981] and Bloesch [1951] showed that the principal also holds for the red-figured period. Thus, in conjunction with the stylistic development of a painter, an analysis of the vases may also provide valuable clues as to the internal chronology of the painter. Thus, on average, if one took two vases of the same type, the chances are that the more slender vessel is later. However, the aesthetic sensibility of the potter himself is also a factor in the slenderness of a vessel. It would stand to reason then, that if the vase constructor were the same, that he were consistent in the way he fashioned his vases, and he were in tune with the aesthetics of the day, then over the course of his career, one should expect, in general, that the stouter vessels would represent the earlier output and slenderer vessels the later.

5.1.2 The Princeton Painter and Potter

In the case of the Princeton Painter, the issue of whether the painter decorated his own vessels or painted the vessels of a single potter may be of some significance in both the attribution of the Princeton Painter and the attempt to establish an internal chronology of the painter's works. The difficulty of the Princeton painter's internal chronology has been pointed out by von Bothmer who remarks "...some painters do not develop in a straight line with clearly recogniseable landmarks, and others show a deplorable tendancy to be inordinately inconsistent in the quality of their works." [Chamay and von Bothmer, 1987] and then proceeds to illustrate how this applies to the Princeton Painter's corpus. An important claim is made by von Bothmer to justify this: that no two vases in the extant corpus of the Princeton Painter appear to have been constructed by the same potter.

There are three reasons why the question of whether the Princeton Painter painted for a single potter is significant. First, if the painter exclusively decorated the work of a single painter then an analysis of the shape of his pots could potentially be used as an attribution tool. Secondly, it is difficult to reconcile the ordering of vases from stout to slender with any meaningful stylistic development in this painter's corpus. Therefore, if indeed the Princeton Painter did pot his own vessels then one would need to account for this discrepancy. Thirdly, if the Princeton Painter exclusively decorated the work of one potter then we may conjecture that potter and painter are the same person and, if this is the case, then the signature "Werxekleides *epoiesen*" left on a fragment in Samothrace[Moore, 1975] leaves some tantalising evidence as to his identity. The term may not be Attic and it is possible that the potter was a metic [Boegehold, 1983, Boardman, 1983].[2] Even if painter and potter are not the same, the fact of their association is itself interesting. Furthermore, together with the *epoiesen* signature, there is a kalos inscription[3] which means it is possible that either potter or painter were literate. This is discussed in more detail in A.2.

5.2 The Problem

The question then arises, can a computer be trained to distinguish between different potters. If so, the preceeding arguments suggest that such a machine could play a useful role in a computer aided attribution of vase-paintings. Furthermore, if such a classifier can be constructed, can it be used to answer the question of whether the Princeton Painter decorated exclusively for a single potter. The aim of the study is both to present an ensemble classifier that can recognise the potter of a vase from a silhouette image of the vase and to apply this classifier to the claim that the Princeton Painter's corpus does not include two pots constructed by the same potter.

5.2.1 Outline of Method

In order to answer these questions two experiments are conducted, one to estimate the performance of the classifier designed to distinguish between potters, and one to test von Bothmer's claim that the Princeton Painter did not paint vessels by a single potter. The experiments are based initially on a majority vote of 75 base classification rules, each obtained by trained LDA on one of 75 different feature spaces. These are described in detail in 5.4. These are applied to the data using the leave one out method (LOO) resulting in 75 sets of predictions for the class-membership of each of the vases. These predictions are combined using an unweighted majority vote.

5.2.2 Experiment One

The aim of the first experiment is to test the performance of the classifier outlined above in distinguishing between the vases constructed by three different potters whose identities are known with reasonable certainty. An ensemble of 75 different classifiers, as described above, is trained on data extracted from 10 vases each of 3 Attic black-figure vase makers: Exekias, Andokides and

[2] Metics were foreign workers in Athens who were technically free (i.e. not slaves) but did not enjoy citizen's rights.

[3] *Onetorides kalos* "Onetorides is handsome" Bonn 365 **R10**.

Amasis. The training and prediction is done on the basis of the leave-one-out method described in 2.5.2. The final set of class predictions comprises a majority vote of the predictions of each of the component classifiers. The true error rate is assumed to be beta distributed according to the method described in 2.5.1 and on this basis both the error estimate and the 95% upper bound on the error are reported. This serves as proof of the concept that a classifier may be trained to recognise the constructor of a vessel.

5.2.3 Experiment Two

The second experiment aims to test von Bothmer's hypothesis that the Princeton Painter painter's corpus contains no two vases by the same potter. If we state the null-hypothesis as "In the corpus of the Princeton Painter, no two vases are constructed by the same potter", then the following consequence should follow if this hypothesis were true: a classifier trained to distinguish between vases constructed by different potters should not be able to distinguish between the works of the Princeton Painter and those of a random selection of vases. To test von Bothmer's hypothesis, an ensemble of classifiers selected from experiment one is tested on all 3 sets of pots with known potters, a set of 10 vases painted by the Princeton Painter, and two control groups of 10 randomly selected vases each, all contemporary to the Princeton Painter.

The reasoning behind the experiment is as follows. An implicit assumption is that the 75 component classifiers in experiment one comprises both weak and strong learners at the task of identifying the work of potters in the set of examples. Of the strong learners it is assumed that most are strong because they are good at distinguishing between the work of potters in general, rather than because they are particularly suited to distinguishing between the specific potters chosen for this study. While there can be no principled justification for this assumption, it is predicated on the fact that choice of the component classifiers was totally independent of the choice of potters. If these assumptions are correct then an ensemble of all the classifiers that performed significantly above chance in experiment one would be expected to predict, with better than chance accuracy, the identity of the potter of a random vase provided it had been trained on examples of that potter's work. Furthermore, such a classifier, under the null hypothesis, should not be capable of distinguishing between the work of the Princeton Painter and the elements of a random set of vases made by contemporaries of the Princeton Painter. Furthermore, under any circumstances, the classifier should not be able to distinguish between the elements of two sets of random vases sampled from the same population using the same distribution.

5.2.3.1 Selecting component classifiers

The question of how to select base classifiers for the ensemble is not simple. The method used here is to base the selection on the performance in experiment one, but the issue of what criteria for selection should be specified requires some reflection. The argument could be made that, in order to get classifiers that make independent errors one should select weaker base classifiers in an ensemble in the belief that they are less stable. However, the circumstances of this particular project are such that it may make more sense to select strong classifiers, but not to make the selection criterion too rigid. First, the diversity of the classifiers is not sought in the learning algorithm used, but in the diversity of the feature spaces. Therefore, poor performance may reflect that the specific feature space is a poor representation for discrimination between classes, rather than that the classifier is unstable. Secondly, from 5.4.1 it is clear that there are families of primary feature extractors that use the same 2D "base" feature extraction technique, but then reduce this to 1D functions using different methods (for example, 4 different $\mathcal{H}$ transforms are defined each using different statistics of the Hough Transform). This means that if a single "base" feature extractor results in 4 poor classifiers, selecting on the basis of weak results may actually lead to correlated errors, rather than diverse errors. Of course a similar argument holds for selecting only the strongest performers, since these may represent all feature spaces defined on a good "base" feature space. Nevertheless, if the criterion for selection is quite strict, then the benefit of having strong classifiers in the ensemble may outweigh the consequences of the correlation. On the other hand, the risk of correlation is too high to justify selecting weak classifiers. For this reason, the approach used here is to select strong classifiers that we are sure to perform significantly better than chance.

Two approaches to selecting strong classifiers from experiment one are either to specify an arbitrary number, say k, and select the k highest scoring classifiers, or to specify an arbitrary performance standard (say a minimum accuracy) and to select only those classifiers that met or exceeded this standard in experiment one. However, to avoid any doubts that either k or the arbitrary performance standard were specifically chosen so that the experiment produced an interesting results, a method with no free parameter was sought.

An intuitive method was chosen instead which is not entirely principalled, but does not allow for any "fine tuning" to select the most interesting results. This is to treat the selection procedure as a hypothesis testing exercise in which classifiers are sought that, with high certainty, perform better than chance. The traditional approach to point null-hypothesis testing is to determine a suitable statistic p, which represents the likelihood of obtaining, under the null hypothesis, data as or more extreme than those observed. This smaller the value of p the greater the evidence against the null hypothesis. Then a preset value, usually $\alpha = 0.05$ is set and p-values lower than this are assumed to be sufficient evidence to reject the null hypothesis. A classifier's vote is included in the ensemble if the classifier's error rate is better than this value at the $\alpha = 0.05$ significance level. To provide extra protection against false positives, correction is made for multiple hypotheses.

5.2.3.2 Correcting for Multiple Hypotheses

When multiple hypotheses are being tested there is the risk of failing to reject the null-hypotheses which grows as the number of simultaneous hypothesis is tested. The naiv̈e approach to solving this problem is to use Bonferroni's correction which involves dividing the α-value by the number of tests which often reduces the required p-scores to such an unacceptably low value that only a handful of experiments pass the test. There are a number of reasons why such a conservative approach is inappropriate in many circumstances but the most important for the present experiment is that while it is good at reducing the false positive rate this is at the expense of the statistical power since it has a high probability of rejecting true positives. In many applications, such as in the majority vote case, the false negative rate is also important as it can affect the diversity of the ensemble.

A popular alternative to the standard Bonferroni correction is Holm [1979]'s modification, sometimes called the Bonferroni-Holm step-down procedure. The idea behind this is to control the false positive rate while also maintaining a low false negative rate. If there are N hypotheses and one has specified an acceptable false positive rate (say $\alpha = 0.05$) the Bonferroni-Holm correction proceeds as follows:

1. The respective null-hypotheses are ordered from lowest p-value to highest.

2. set $k = 0$. The lowest p-value is compared with $\frac{\alpha}{N}$ and if it is lower it is rejected, k is incremented and the procedure continues, else the procedure stops

3. the next lowest p-value is compared with $\frac{\alpha}{N-k}$ and if it is lower, the null-hypothesis is rejected, k is incremented and the procedure continues, else the procedure stops.

4. The previous step is repeated until a p-value is greater than $\frac{\alpha}{N-k}$ at which point all null hypotheses that have not been rejected are accepted as significant.

In this section, the Bonferroni-Holm method is used to select the classifiers that perform better than chance. Since there are so many classifiers, using the strict Bonferroni method (i.e. dividing α by 75) has the danger of reducing the statistical power of the test too greatly. On the other hand, if only a few weak classifiers are included in the ensemble, their errors are expected to wash out in the majority vote. Therefore, for the purposes of this study the less strict Bonferroni-Holm method is sufficient.

Because of the small number of samples and in order to maintain consistancy, the p statistic is defined on the posterior error estimate (explained in 2.5.1) rather than the traditional likelihood of the null-hypothesis. While there is no coherent theoretical justification, it is simply an intuitively appealing method of selecting "good" classifiers, and it is self evident that it will do the job. First, the value p_p is defined as follows

$$p_p = 1 - \int_0^{\varepsilon_0} p(\varepsilon) d\varepsilon \tag{5.1}$$

where ε_0 is the error rate expected by chance and $p(\varepsilon)$ is the probability density of the true error (calculated according to the method established in 2.5.1 and treated, in the Bayesian manner as a variable rather than a fixed quantity). A visual interpretation of this statistic is provided in figure 5.1. p_p, like the traditional p value serves as evidence against for null hypothesis - the higher p_p, the less justification for rejecting it. On a three way classification, an error rate of $\frac{2}{3}$ is expected by chance and this is ε_0. By tradition, the typical $\alpha = 0.05$ global significance level is chosen.

5.3 Data

The data are composed of silhouette images of pots constructed by different potters. The Athenians built vessels according to a number of predefined shapes to which examples usually conformed very closely. To make comparisons meaningful, the same general shape has been chosen for the data set: the belly amphora. The belly amphora had a belly constructed in a continuous curve surmounted on a foot that could either be in two stages or inverse echinus. The shape tapers towards the neck and has a mouth that is usually echinus or flaring. The handles are thrown separately from the rest of the vase and these are oriented horizontally, from a point high on the

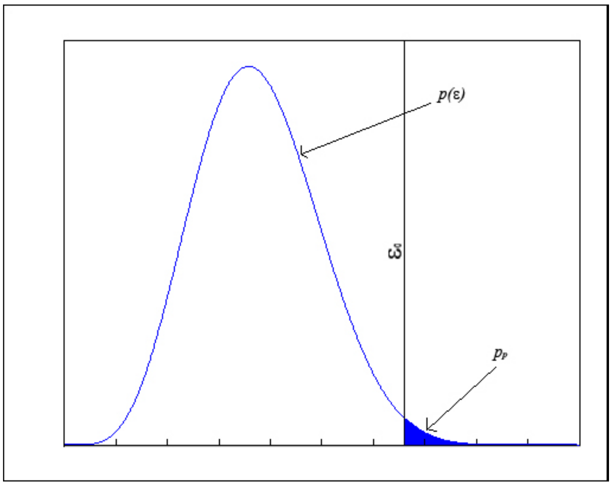


Figure 5.1: A visual interpretation of the p_p statistic.

belly or low on the shoulder and terminating on the neck, close to the lip. The term belly amphora indicates that the amphora's whole body is thrown in one piece and essentially is a large belly with undifferentiated neck and shoulder. The belly amphorae in figures 5.3 and 5.4 may be compared with a neck amphora in figure 5.2 in which the neck is thrown separately from the body. The reason for the choice of the belly amphora is that the Princeton Painter decorated vessels of this shape far more often than any other shape.

5.3.1 The Agents

There are a number of potters whose work may distinguished either on the basis of their distinctive style or because they have signed their vessels or both. However, in order to select appropriate potters for the validation a couple of conditions apply. The first is that the potters should be contemporary with the Princeton Painter as we wish to attenuate the effect of chronology on vase-shape. A second is that at least some of their vases are signed, and that the styles of the rest are distinctive enough that they are considered by most scholars to be the work of one potter. Finally, they need to have some medium sized belly amphorae in their corpora, since the Princeton Painter mostly decorated this type of vessel. Of those that meet the criteria, the clearest cases are Amasis, Exekias and Nikosthenes. However, Nikosthenes is not suitable for this study as he is a mannerist whose style is so exaggerated that his amphorae are considered to belong in a separate class from those his contemporaries. Therefore, the three potters from whose corpora

Figure 5.2: A Neck-amphora in which the juncture between the shoulder and the neck is articulated.

Figure 5.3: The type B amphora (from c. 600 BCE).

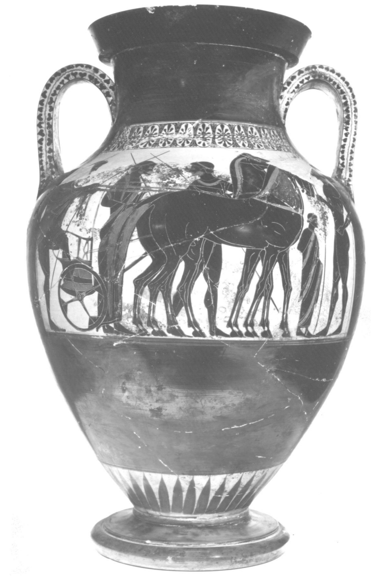


Figure 5.4: The type A amphora (from c. 550-540 BCE).

the validation sets have been drawn are Amasis, Exekias and Andokides.

5.3.1.1 Exekias (10 vases)

Exekias has already been introduced and discussed, both in 3.4.6.2 in respect of his forms, and in 4.3.1.4 in respect of his sense of balance and proportion. Exekias was also a potter who signed a number of his vessels. The pottery of Exekias reveals to some extent his debt to Group E but includes innovations that distinguish him from these groups. One innovation that is particularly relevant here is the introduction of the type-A amphora which Exekias uses frequently. The major differences between type A and type B, the latter of which had been the standard for much of the 6th century, are visible in figures 5.4 and 5.3. The type A amphora had a foot in two stages rather than inverse echinus, and the handles are often rectangular rather than circular in section. Exekias' belly amphorae are mostly of type A (there are only two type B in this data set). By contrast, Amasis and the Princeton Painter's belly amphorae are almost all of type B.

5.3.1.2 Andokides (9 vases)

The name Andokides appears 11 times, 9 times as *epoiesen* signature, once on a kalos inscription dated to around 540[4] and his name, together with Mnesiades, is signed on the bronze support of a dedication found on the Akropolis (the dedication itself is lost). His vases are traditionally dated from 540 to around 510 and it is in his workshop that some scholars believe the red-figure technique was born. While there have, as yet, been no monographs on the work of Andokides, his work has been analysed in some detail because of his importance as the first recognised red-figure potter.[5] The selection chosen from Andokides includes one type B amphora, but the rest are type A.

5.3.1.3 Amasis (10 vases)

Amasis is the name signed as maker on at least 9 extant vases and the style of the painting that decorates these pots was very distinctive. The painter, who appeared to paint exclusively the pots of Amasis,[6] is therefore referred to as the Amasis painter. Because of the distinctive painting and potting, the Amasis painter had long been known as an artistic personality by the time Beazley described his style in *Dev* and expanded on this in Beazley [1931]. Since then, there have been many publications on aspects of his style including a monograph by Karouzou [1956]. To celebrate the centenary of Beazley's birth, a travelling exhibition of vases by Amasis/The Amasis Painter took place in the US in 1985 out of which two books were published [von Bothmer, 1985, Robertson et al., 1987]. The former was a publication of all the pieces in the exhibition by von Bothmer and included a thorough analysis of Amasis' pottery style and some attributions based on profile analysis. Amasis is also responsible for some interesting questions and problems in Attic black-figure which are beyond the scope of this dissertation.[7] Equally distinctive is the style of the potter himself which has been discussed at length by von Bothmer [1985].

[4] *Andokides kalos dokei* "Andokides appears beautiful" appears on a black-figure hydria by Timagoras, Louvre F38 *ABV* 174.7, 667 *Para* 72 *Add* ^{2} 49

[5] Andokides signed as potter of black- and red-figure pots as well as bilingual amphorae - that is pots painted with different techniques on either side

[6] But not the other way around. Amasis has almost certainly potted for other painters, including two paintings by Lydos

[7] The most important of which is surely that Amasis has always been assumed to be named after the Egyptian king A-Ahmos who became a philhellene in 569 [Cook, 1948] suggesting the potter must have been born after this date. But the dating of his style appears to be from 555. Boardman [1958] counters that he could have been Egyptian by birth, and not necessarily Athenian.

5.3.1.4 Random Selection (20 vases)

In addition to the three known potters, two control groups of random vases are included. These are roughly contemporary vases and are exclusively type B amphorae to match the Princeton Painter's corpus. A number of factors complicate the issue of how to ensure that the selection of vases in the control groups are indeed random. Simply sampling randomly from *ABV* is unsuitable because

1. *ABV* includes many different shapes, whereas this study requires only amphorae
2. some potters' styles are so distinctive that their selection might bias such a small sample study
3. the range of dates of vases spanned by *ABV* is much larger than the plausible dates for the Princeton Painter's work
4. many of the vases in *ABV* do not have suitable images available
5. many of the vases in *ABV* are fragments and this study requires whole examples

Therfore the following strategy was used: A random number genarator was used to select from pages in *ABV* that are associated with the period contemporary with the Princeton Painter.[8] This amounts to anything from the painter of Acropolis 606 to the Leagros group. If the painter decorated type B amphora then one of his black-figure amphorae was chosen at random (again using a pseudo-random number generator). If that amphora does not have an image in *CVA*, the closest type B amphora by that painter (in terms of Beazley's numbering) for which there exists *CVA* images was chosen instead. Otherwise the process was repeated.

5.3.2 Profile Analysis

The analysis of pottery shape has traditionally been conducted on the vertical profiles of the vases. The profile is usually a drawing representing the perpendicular distance from the vertical medial axis to the outside of the vase

[8] *ABV* is roughly chronological, but not strictly so since Beazley chose to place cognate groups of painters together even if they weren't exactly contemporary. So, for example, most of Group E is earlier than Exekias, but he is certainly related - so they go together, whereas the Painter of Berlin 1686 probably starts earlier than Exekias, but Beazley puts him considerably further in the book because he is lumped together with other insignificant painters like the Princeton Painter.

as a function of the vase's height. Exactly how one defines the two dimensional distance metric on a three dimensional object is not been explicitly mentioned by Classical archaeologists. In general, perfect symmetry of the vase is assumed and the measurements are made from the medial axis to the outside in one direction only (usually at ninety degrees azimuth from the handles). Of particular importance to most scholars are the profiles of the neck and the foot, since these may not be easily recovered from photographs. An example of a vase attributed by Beazley to the Manner of the Princeton Painter is provided in figure 5.5. The methodology of making inferences from profiles is often referred to as profile analysis and the approach adopted in this study is based to some degree on this. Before explaining how such profiles are extracted from the images themselves, a brief history of the previous techniques is discussed and a motivation provided for the method used in this study.

When Bloesch [1940] published his monograph on potter attribution of cups, he introduced two methods that would have considerable impact on the way in which vases were published. First, he pointed out that the way in which photographs of vases had been taken[9] was unsuitable for a thorough analysis of the shape because of the effect of distortion of the curvature of the vase's profile. Instead, he established a method for taking these photographs by which the lens of the camera was at the same level as the greatest diameter of the vase, and that the distance between lense and object was at least 6 times the largest dimension of the object. This formula has been taken to heart in most subsequent publications of artifacts, meaning that many photographs published since do meet the requirements for analysing the shape of the vase.

The second method introduced by Bloesch was the drawing of profiles of the necks and feet of vases, since these were areas in which the subtleties of the shape might not be clearly discernable from photographs. Since then, it has become an established part of the practice of publishing vases to supply drawings of the vase profiles. Although these profiles are generally quite accurate, the scholars responsible for these drawings generally do not explain how they are derived, and different methods are employed by different scholars, with the specific methods used often relying on the ingenuity of the scholars themselves.

Mackay [1985, 1981] proposed a mechanical method of generating consistent profiles from vases. The method consisted in using a simple device to take accurate measures of the width of the vase at fixed intervals along

[9] Often with the lense very close to the vase from an angle that would best capture the designs drawn on the surface.

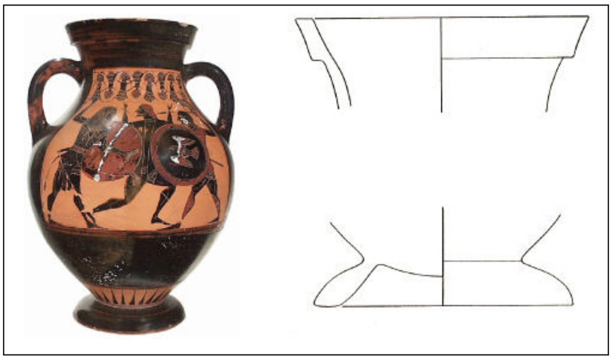


Figure 5.5: An example of a vase (Oxford 1965.131) and its neck and lip profiles from the *CVA* Oxford 3 Great Britain fasc. 14.

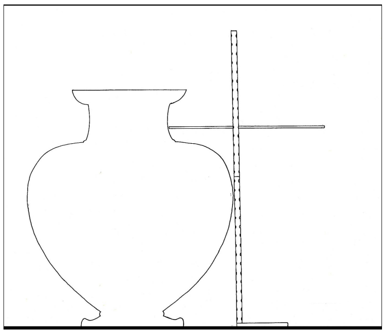


Figure 5.6: Mackay's device for measuring vase profiles from Mackay [1981].

its height. The device is composed of a vertical metal frame with a series of horizontally extendable rods at fixed intervals along the vertical axis (figure 5.6). The vase is placed so that it touches the metal device, and the rods are extended till they touch the surface of the vase. The amount by which the rods are extended from the vertical metal frame is measured and from this, points along the profile may be sampled. The profile diagram (for example figure 5.7) is created by plotting the points on graph paper and smoothly joining (interpolating) the points, except of course, where there are disjunctions in the vase's curvature such as at the neck and any articulations. Mackay has used profile analysis to good effect in establishing a chronology of Exekias' works [Mackay, 1981]. Unlike many scholars, Mackay pays particular attention to the curve of the belly of the amphorae in attributions and chronologies, though she acknowledges the importance of neck and lip profiles [Mackay, 1985, p.236].

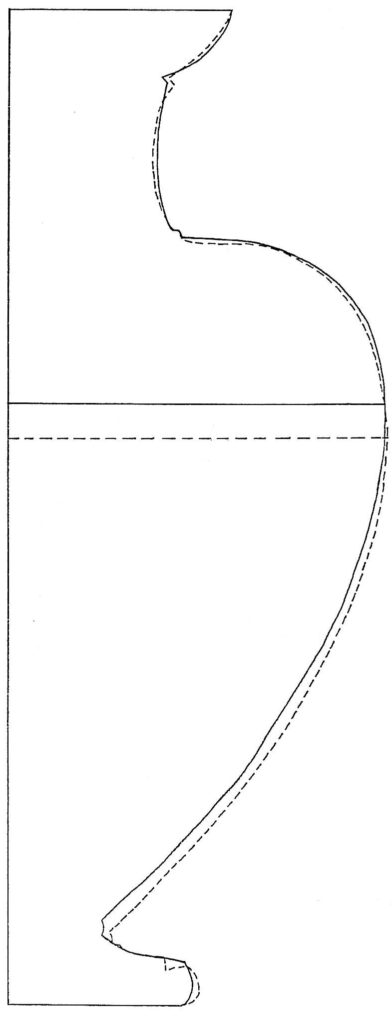


Figure 5.7: Superimposed vase profiles comparing two Exekian vases - London B210 and Dublin 1921.97[Mackay, 1981, p.26 fig.3].

While Mackay's method of extracting profiles is superior to those used before, it also requires *ex vitrine* examinations of the vases and this is laborious and more importantly, very expensive and certainly out of the reach of some scholars. Instead, profiles may be extracted from photographs using a combination of automatic and manual image processing so the technique may still be applied even without *ex vitrine* examinations of the vases themselves. This study is, however, limited by the fact that without access to the vases themselves, no precise study of the feet and lips of the vases can be conducted - these features are not faithfully reproduced in photographs. Nevertheless, a similar method has been applied with considerably success by Durham on modern Cypriot vases and this provides some hope that there is some potential in the technique. Although Durham was unsuccessful in replicating this effect on black figure, there are sufficient differences between his study and this to justify another trial. First, Durham was attempting to distinguish between vases painted by the Swing Painter and by the Antimines painter by examining the shape of the vases, and it is not widely held that either the Swing painter or the Antimines painter's corpora were the output of individual potters. Secondly, Durham used a single shape description algorithm, the GHT, which may not have been sensitive enough to subtle differences in shape - the Attic tradition was very rigid.

5.3.3 Image Pre-Processing

The goal of the image preprocessing was to extract a 400 pixel high image of the vase in silhouette in which the vase is black against a white background. Because the images of the vases were taken under varying conditions, manual image processing was required. The same technique was followed from vase to vase. The processing was carried out in Photoshop CS 2 and many of the image processing algorithms are proprietary and therfore an exact mathematical description of the process cannot be provided. However, the method was applied consistently from vase to vase and is described below.

Vases were, in the first instance, taken from pictures from the *CVA* where available online and from the author's own corpus where these were not available. Even though these were not taken by the same photographer, for the most part, Bloesch [1940]'s criteria were used in regard of distance from lens and height of camera. Furthermore, for the most part, the author's own collection is composed largely of images taken directly or indirectly from the *CVA*s. The preprocessing of the images was done as follows. First, because even though the photographs have been taken at a substantial distance from the vase the curvature of the foot and the lip in the horizontal plane project slightly into the image in the vertical plane as illustrated in figure . The first

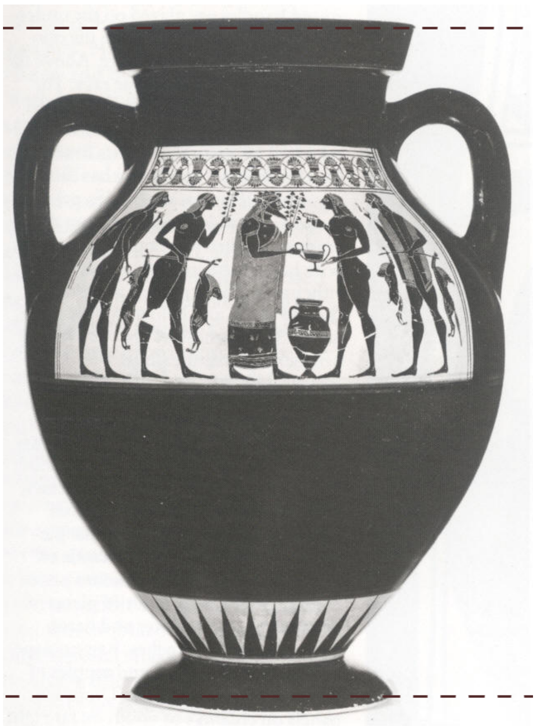


Figure 5.8: A belly amphora photographed from the front, illustrating the "bulge" where the curvature of the foot and the mouth project into the image plane.

step of the preprocessing is to remove this by cropping the image. The image is then scaled so that it is 650 pixels high. The lasso function is used to select contiguous areas of the background and remove them. Whatever background is not selected by the lasso function is removed manually using the selection mask function together with the pencil and the eraser tools. In most cases, very little correction is required as the lasso function is capable of removing most of the background. Despite due care being observed, such manual processing leads to artifacts which may be recognised by the ensemble, and therefore the edge of the vase is smoothed. In order to achieve the smoothing effect, the image is blurred using a gaussian kernel with $\sigma = 3$ pixels. The vase is then thresholded so that the new edges match the edges of the original image as closely as possible - the comparison is conducted by subtraction of the silhouette from the original.

5.4 Feature Extraction/Model Selection

One of the key issues in this study is to find the optimal classifier or an optimal set of classifiers that are capable of distinguishing between potters. In fact, by the No Free Lunch theorem we are not guaranteed that any specific classification scheme will perform better than chance without some prior knowledge of the problem. We assume that a set of classifiers exists but that we do not know a formal specification of that set that will allow us to explicitly sample from it. Since we use the same classifiers trained on different features, this problem reduces to that of choosing a feature space that is optimal for linear classification. However, the Ugly Duckling theorem means that, again, without using prior knowledge, no feature space should be expected to be better than any other. A search through the literature is unrevealing in itself because most of the shape descriptors defined are domain specific and therefore cannot be guaranteed to work on a domain different from the one for which they were intended. While this does not preclude any one of these descriptors from being optimal, there is no principalled method to choose between them.

There are at least two possible solutions to the problem, both of which involve casting the net very wide by creating a large superset of possible feature extractors and subsampling from this set. First, this may be achieved by holding out some of the data and using this to find an optimal feature set for a particular classifier type. The second is to use classifiers trained on all of the feature spaces and combine their predictions using an ensemble method. The idea is that diversity of classifiers will be achieved because most of the features are chosen independently of each other. In the first experiment, the second method is used by itself to establish a baseline error for the system and to provide proof of concept. A combination of the first and second methods is used to select appropriate features for experiment two. The method involves selecting from the classifiers trained on data represented in each of the 75 feature spaces, a smaller subset that can, with high confidence, be thought to perform better than chance.

5.4.1 Primary Extractors

The main primary extractors have already been discussed in 3.4.7. However, the analysis of quadrant statistics has been dropped, and the $\mathcal{R}$-transform and Moving Statistics (5.4.1.1) added to the ensemble. Unlike Chapter 3 where the aim was to keep the size of the feature space low, here a large set of initial feature spaces is preferable, and this set will be pruned by the method described in 5.2.3. To this end, the $\mathcal{H}$-Transform is modified using certain point statistics - in particular the central moments, to create 4 different feature extractors. Before describing the specific modifications, the central moments themselves will be defined.

The kth central moment μ_k of a random variable X is $E[X - E[X]]^k$. The first central moment is proportional to the mean and the second central moment is the variance. Defined on the 3rd and 4th moments are the higher order statistics skewness (γ) and kurtosis (β_2) respectively:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \tag{5.2}$$

$$\gamma = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} \tag{5.3}$$

The skewness measure is sensitive to asymmetries in the distribution around the mean, whereas kurtosis is sensitive to peaks in the distribution. The kurtosis of the normal distribution is 3 so often this is subtracted from the kurtosis and reported as the excess kurtosis ($\gamma_2 = \gamma - 3$).

5.4.1.1 Moving Statistics

This is a simple shape descriptor that is inspired by the sliding window methods used in some hand-writing recognition systems. A $10 \times w$ pixel window is translated vertically along the silhouette of the vase. At each point along the translation, the statistics for the pixels in that window are recorded. This results in 4 different feature vectors for each silhouette, 1 each for mean, standard deviation, skewness and excess kurtosis.

5.4.1.2 The $\mathcal{H}$-Transform

In 3.4.7.3 a new shape descriptor was defined on the Hough transform, called the $\mathcal{H}$-transform. This assigns to every angle in the accumulator the mean of the different values of θ as follows

$$H_a(\theta) = \frac{1}{P}\sum_{\rho=0}^{P} M_{\theta,\rho} \tag{5.4}$$

$$= \text{mean}_\rho(M) \tag{5.5}$$

where M is the accumulator array for the Hough transform (defined in chapter 3.4.7.3) and mean_ρ is the mean with respect to ρ. In addition, 4 different feature vectors may be defined on M, one for each of the statistical moments. For example, instead of the mean, we define $H_{sk} = \text{skewness}_\rho(M)$ Analogously H_s represents the standard deviation and H_k the kurtosis. Thus from the $\mathcal{H}$-Transform, 4 different feature extractors are defined.[10]

5.4.1.3 $\mathcal{R}$ - Transform

Tabbone et al. [2006] introduced the $\mathcal{R}$-transform which is a descriptor based on the discrete Radon transform of an image. While the $\mathcal{R}$-transform is not limited to skeletonised images or curve descriptions, its effectiveness on such problems when used together with LDA has been demonstrated by Barrat and Tabbone [2007] who use this combination for symbol recognition.

The Radon transform is a generalisation of the vertical and horizontal projections to arbitrary angles. Given a set of lines defined on a binary image **BW** described in parametric form $L = x\cos\theta + y\sin\theta$, the radon transform maps a binary image onto a matrix T_R in which the columns represent indexed values of ρ and rows represent indexed values of θ. In the case of a continuous image, I, the Radon transform is defined by

$$T_R(\theta,\rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,y)\delta(x\cos\theta + y\sin\theta - \rho)dydx \tag{5.6}$$

where δ is the Dirac delta function. In discrete terms this translates to

$$T_R(\theta,\rho) = \sum_{x=0}^{X}\sum_{y=0}^{Y} \mathbf{BW}(y,x)\delta(x\cos\theta + y\sin\theta - \rho) \tag{5.7}$$

[10] although their results may be correlated.

where δ represents the Kronecker delta function and X and Y represent the image x and y dimensions respectively.

The $\mathcal{R}$ transform is simply the sum of the Radon transform over different values of θ or

$$\mathcal{R}(\theta) = \sum_{\rho=0}^{P} T_R(\rho, \theta) \tag{5.8}$$

where P is the greatest possible value of ρ. The $\mathcal{R}$ transform is translation invariant, a rotation of the image results in a translation of θ in the $\mathcal{R}$-transform, and changes in scale produce a change in the scale of the amplitude. This means that the $\mathcal{R}$-transform may be made scale invariant by dividing the transform by its mean, and the histogram of the $\mathcal{R}$-transform is rotation invariant.

Implementing the $\mathcal{R}$-transform may thus be achieved by replacing the summation in (5.8) with mean_ρ giving $\mathcal{R}_a(\theta) = \text{mean}_\rho(T_R(\rho, \theta))$. In addition, by exchanging the mean with the standard deviation and higher order statistics, we may define 3 further feature spaces on the Radon transform in an analogous manner with the $\mathcal{H}$-Transform.

5.4.1.4 Secondary Feature Extraction

In addition to PCA70/90, four different secondary feature extractors are used in this study. The first two are simply the first and second derivatives of the feature vectors which are estimated by subtracting each element from the next element with no smoothing. The result is that the derivative is one element shorter than the original feature space, and the second derivative two elements shorter. The third method is to use the first 10 components of the discrete cosine transform (DCT) of the feature vector. The DCT transforms the vector from amplitude-distance space to frequency-power space. One way of approximating a discrete time signal (such as a feature vectors) is as a linear series of N trigonometrical functions, such as sine or cosine waves, each weighted by a suitable real number. The DCT is one method used to decompose the signal into its components.

There are 8 standard forms for the DCT (each marked by numerals I-VIII) but the most popular, the DCT-II, is often simply called the DCT Ahmed and Rao [1975]:

$$DCT[f(n)] = 2\sum_{n=0}^{N-1} f(n)cos(\omega_k(n + \frac{1}{2})) \tag{5.9}$$

where $\omega_k = \frac{k\pi}{N}$ (for $k = \ldots N$). For the purposes of this study, the MATLAB function dct(.) has been used to implement the DCT. This is similar to the version above but with a slight modification:

$$DCT[f(n)] = w(k)\sum_{n=1}^{N} f(n)cos\frac{\pi(2n-1)(k-1)}{2N}, k = 1\ldots N \qquad (5.10)$$

Here $w(k) = 1/\sqrt{N}$ when $k = 1$ and $\sqrt{2/N}$ otherwise.

Finally, the fourth secondary feature extractor takes as its components the 4 statistics based on the central moments of the original feature space. In all cases, PCA90/70 is still used in addition to the other secondary extractors to ensure that the covariances matrices are non-singular.

We therefore have the following permutations for feature extractors

1. $\mathcal{H}$-Transform or $\mathcal{R}$ transform or Moving Statistics (12 variants) + secondary extractor (3 variants) + PCA = 36 feature spaces.

2. Curve signature or Vertical Projection or Horizontal Projection (3 variants) + secondary extractor (3 variants) + PCA = 9 feature spaces

3. primary extractor (15 variants) + PCA = 15 feature spaces

4. primary extractor (15 variants) + central moments + PCA = 15 feature spaces

This results in a total of 75 feature spaces tested. The entire classifier design is illustrated in figure 5.10

5.5 Results

5.5.1 Experiment 1

The baseline error rate of the complete ensemble of classifiers on the 29 vases with known constructors is 0.089 with a 95% upper bound of 0.23. Given that the expected error due to chance for a three-way classification is 0.677, this results suggests strongly that the approach adopted in this study is capable of distinguishing between pots by different potters. Table 5.1 is the confusion matrix for the entire set of classifiers. Interestingly, it is clear that even without selecting for the best classifiers, the ensemble is capable of discerning the Princeton Painter's vases from those of the control groups.

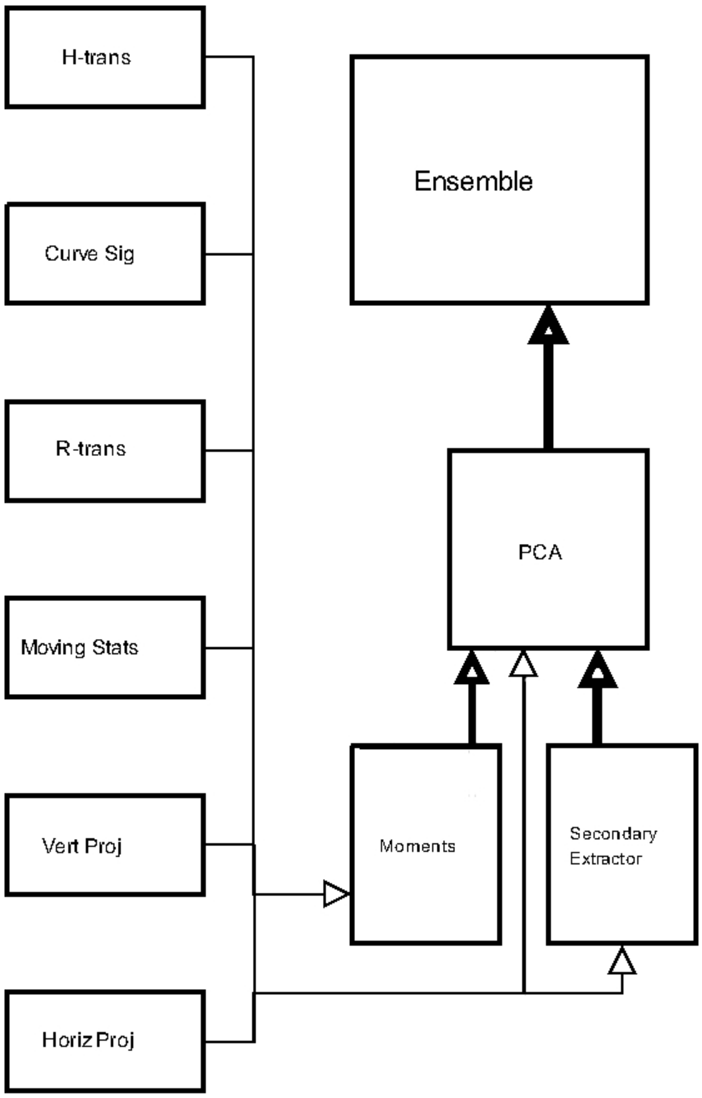


Figure 5.9: The classifier design schematic illustrating the role of primary and secondary feature extractors in the overall ensemble.

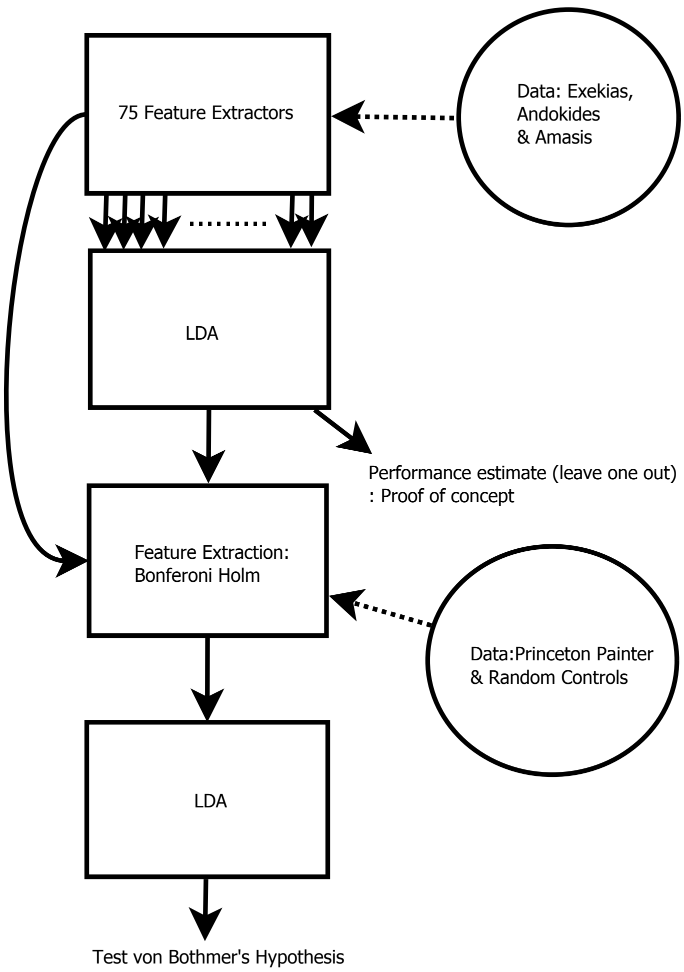


Figure 5.10: The classifier design schematic illustrating the learning process.

	Prince	Amasis	Exek	Andok	Contr 1	Contr 2
Princeton	8	0	0	0	2	0
Amasis	0	10	0	0	0	0
Exekias	0	0	8	0	0	0
Andokides	2	0	1	8	0	0
Control 1	0	0	1	1	4	5
Control 2	0	0	0	0	4	5

Table 5.1: Confusion matrix for ensemble classification based on all 75 feature spaces

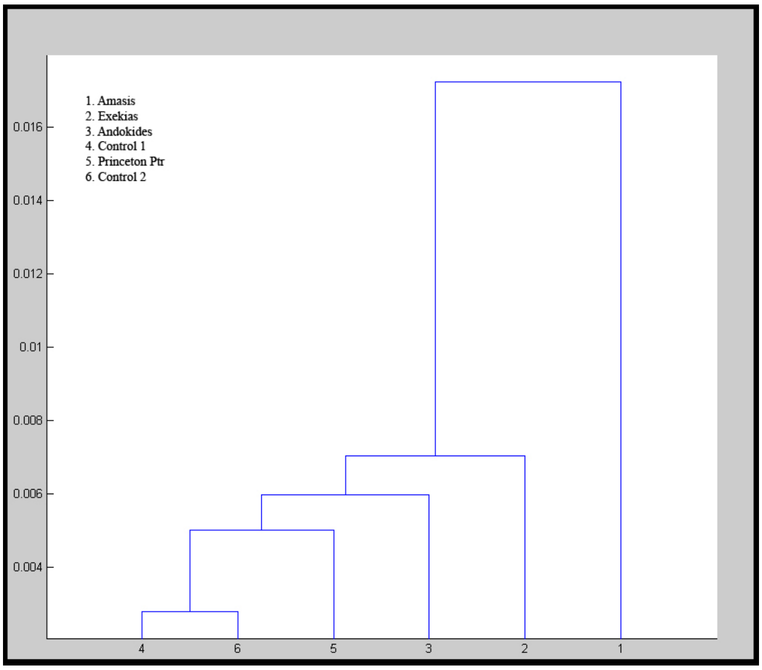
	Prince	Amasis	Exek	Andok	Contr 1	Contr 2
Princeton	8	0	0	0	1	0
Amasis	0	10	0	0	0	0
Exekias	0	0	8	0	0	0
Andokides	2	0	1	9	0	0
Control 1	0	0	1	0	5	5
Control 2	0	0	0	0	4	5

Table 5.2: Confusion matrix for ensemble classification based on an ensemble of classifiers that performed better than chance on the 'known constructor' set

5.5.2 Experiment 2

Figure 5.2 shows the confusion matrix genarated by the classifiers that scored greater than chance on the set of known vase-constructors, correcting for multiple comparisons using the variant of the Bonferroni-Holm stepdown procedure described in 5.2.3.2. Of the 10 vases by the Princeton Painter, the ensemble correctly predicted 8 of these, giving an error rate of 0.2 with a 95% upper bound of 0.48. Again, even with the very strict assumption that if the null-hypothesis were correct, such a strong learner would only misclassify Princeton Painter vases to either of the random control groups, we should expect an error rate of 0.667 under the null-hypothesis. This may mean that at least some of the Princeton Painter's vases were constructed by a single potter. The sanity check that the machine should not be able to distinguish between members of the two control groups is also validated by both the ensemble of all classifiers (table 5.1) as well as by the ensemble of strong classifiers (table 5.2).

5.5.2.1 Reconsidering von Bothmer's Hypothesis



(a)

Figure 5.11: Dendrogram distance between potters based on the cumulative confusion matrix.

However, given that the null hypothesis states the opinion of such an eminent scholar, and the most important exponent of potter attribution, this result requires further interrogation. First, we visualise the relationship between the potters based on the predictions of all the 75 classification rules. To do this, in a similar manner to the method employed in chapter 4 to determine stylistic distance, a distance measure was defined on the pooled or cumulative confusion matrix. The confusion matrix is simply the sum of the 75 confusion matrices of all the predictions of the 75 classification rules. To convert the confusion measure into a distance measure, as in chapter 4 (4.2.4.1), if C is the confusion matrix then the distance measure $D(i,j) = \frac{1}{C(i,j)+C(j,i)}$. From these measures, hierarchical agglomerative clustering (single linkage) produces the dendrogram in figure 5.11(a). The results show that the Amasis painter is the most distinctive of all the potters (unsurprising given his distinctive style). Next most distinctive overall is Exekias followed by Andokides. The two random control groups are clustered together and the Princeton Painter is the closet of the known potters to the random group, but he is clearly distinct. Considering the closeness of the Princeton Painter to the random groups, however, it is worth considering conditions under which von Bothmer's hypothesis could hold even given the results presented here.

Careful consideration should be given to the possibility that von Bothmer's hypothesis is true but that an artefact in the experimental process is responsible for the strong result. There are many possibilities for error in sampling and pre-processing (since these were done by hand) despite best efforts to ensure consistancy. One particular possibility that requires discussion is that sampling bias may account, in part, for the strong result. In particular, it is possible that the control groups may be sampled from a different population of potters than those who worked with the Princeton Painter, even if the Princeton Painter's output does not contain any two vases by the same potter. Two possible scenarios are consistent with this: that the control group is biased in favour of better vases or vases by more well known painters, or that the Princeton Painter worked with a small group of potters that shared a similar potting style.

First, the vases were chosen from images that could easily be processed and this certainly could introduce a sample bias that favours better examples of vase-paintings, or examples of vases decorated by better known painters. For example, there may be a tendency on the part of photographers to take more care over the photography of an Exekias than a poorer vase painter. Since the sampling method used here rejected vases if good photographs were not available, this could have biased the random control groups towards better quality vases or vases by better known painters. Furthermore, it is possible that well-funded museums could afford to spend more time on photographs and therefore more likely to publish *CVA*'s. This again would provide a slight bias to the more well known artists whose works are better represented in large collections than in smaller ones, owing to the high prices they command. This would also lead to artefacts, since the Princeton Painter's vases are relatively poor, while those of the control group contain a mixture of poor and good artists.

Despite these arguments, the classifier correctly attributing 8 out of 10 in a 3 way classification is extremely unlikely to occur by chance and even if there were a bias towards better vases in the control group, this would not alone be able to account for such a strong result. More importantly, the control group contains a mixture of good and poor vases and a mixture of well known and obscure painters. It is implausible therefore that sampling bias alone can account for this result.

The second possibility, that the Princeton Painter did not work with a single potter, but a small number of potters, is entirely consistant with the result. There are a couple of possible scenarios that fit this. First, perhaps the Princeton Painter only liked particular kinds of surfaces to paint on, and was therfore particular about the way in which the amphora he worked on were constructed. There is also the possibility that the Princeton Painter worked with a collective of some sort. Beazley identified 4 other stylistic groups that he considered to be closely related to the Princeton Painter. He referred to them collectively as the Princeton Group. The stylistic similarity between these painters may be due to their working closely together or their belonging to a single workshop. If this is the case, it is quite possible that the workshop had a number of potters that worked together. If this is the case then it would not be surprising if their outputs shared some common traits because they would have learned their trade from one another, probably having started as apprentices in these workshops. This scenario is consistant with both the results of experiment two and von Bothmer's hypothesis.

These scenarios are very speculative, however, but they do serve as a word of caution against rejecting von Bothmer's hypothesis outright. Given the small sample size, we cannot state categorically that the experiment has disproven it. Instead it is safer to interpret the result more conservatively as placing strain on von Bothmer's hypothesis but revealing instead that there is something homogeneous about the potting style of the Princeton Painter's vases, but this may not necessarily mean they are constructed by the same person. Furthermore, as it is argued in the next chapter, all results of this sort require replication, and perhaps the question may be answered in subsequent studies using more and better quality data.

5.6 Conclusion

The method presented in this study is capable of recognising the potter responsible for constructing a vase with moderately low error rate. The error rate is not as low as that for the Morellian approach, but unlike the Morellian, the approach presented here does not rely on the art historian first describing the salient features of a painter's style. More importantly, because Beazley did not use shape as a major attribution criterion, it is the closest example in this study of a method capable of revealing the actual identity of a painter, rather than the dissertation's more modest aim of proving that an "auto-Beazley" is possible. An application to the Princeton Painter can provide an example of why this is the case. The Princeton Painter's identity was established by Beazley on the basis of the paintings alone and not on the manner in which the vases were constructed. If the Princeton Painter were not a real artist, just a collection of works whose style appeared coherent to Beazley, then it would be a remarkable co-incidence if all such works were the output of a single potter.

In this regard, von Bothmer hypothesised that the Princeton Painter's extant output is not only not the work of a single potter, but that no two of his extant pots were made by the same man. This hypothesis comes under strain in the light of the results of experiment 2 which shows that an ensemble of classifiers trained to distinguish between known potters does not fail to identify the work of the Princeton Painter as being distinct from that of two random groups of potters. Even if the Princeton Painter's output is not the work of a single potter, there is still something homogeneous about the shape of the pots that Beazley assigned to the Princeton Painter by methods other than an examination of the shape of the pots. As has already been stated, this result puts strain on von Bothmer's hypothesis. However, the result should be interpreted cautiously. Not only could the result be explained by sampling bias, but there is also the possibility that the vases were made by different potters, but that these potters, for some reason, formed a homogenous group, such as a single workshop. In any event, given the constaints imposed by the data, further studies are required before we can accept these results as conclusive.

Chapter 6

Conclusion and Postscript

6.1 Introduction

The present chapter serves both as a conclusion and as an attempt to wrap up certain loose ends that do not fall within the stated aims of the thesis. Instead this chapter deals with issues that contextualise this dissertation within what is a nascent discipline. It is the belief of the author that attribution through pattern recognition and machine learning is an emerging interdisciplinary field that is still in its infancy, and that considerable study is still required before automatic attribution becomes a reality. There are a number of issues that are not addressed in this study, but which will need to be addressed if computer-based attribution methods are to be used in practice. Two particular areas will be the focus of this chapter. First, while this study only presented three approaches to attribution, there are many more possible approaches to designing classifiers for attribution, and this in itself could be a specific area of inquiry. This chapter will suggest a way forward for conducting research into new classifiers, and ways in which these classifiers may be used in combination. Secondly, the study was limited by the availability of data. While photographs are available for almost all of the vases[1] their quality varies a great deal and for the most part, they are in any event unsuitable for the types of analyses undertaken in this study or in studies that require even more precise tools. We believe that changing the methods by which ceramics are published will not only remove this limitation from subsequent studies into attribution and automatic artefact description in archaeology, but will facilitate archaeological research at all levels (i.e school, undergraduate and scholarly research). The second part of the chapter, therefore, surveys some methods by which such archaeological data are published, and makes some suggestions for future research into improving it.

6.2 Further Study into attribution

The stated aim of this dissertation was not an exhaustive survey of all possible methods of attribution of black-figure vase-painting and there were many features of vase-paintings that are used as attribution criteria by connoisseurs that have been ignored in this study. For example, composition, neatness, symmetry, size were ignored, as were representations of women, subsidiary decorations, animals, and inanimate objects. Future studies may well explore these areas as the basis for computer-aided attribution. This section both

[1] A handful of vases have been lost - some to private collectors and others to bombing in the second world war.

explores some of the possibilities and suggests some guidelines for the way in which such research should be conducted.

There is already a growing body of literature that applies pattern recognition and machine learning principles to the study of attribution (although not of Greek art). The study of Attic vase painting would seem naturally to belong to the same area of research, yet none of the methods that appeared useful in the study of post-Renaissance western art have proved to be useful in this archaeological study. Yet intuitively it seems likely that many of the art historical principles that underlie attribution should apply equally to vase paintings as they do to other forms of art since they derive from the same human urge to artistic expression. One of the aims of future research should be to impose some structure on the pattern recognition approach to attribution and establish attribution as a scientific discipline. The ultimate aim of the discipline would be to provide theories that unite these different strands and approaches to attribution in different media.

Three issues are of importance here. First, the way in which we learn from the results of studies such as the present one need to be systematic. Secondly, there should be studies to try and model attribution as more than a statistical process, perhaps based on findings from relevant disciplines such as neuroscience, psychology and cognitive science. And finally, there needs to be some method for using different classifier designs in concert. These issues are dealt with in order.

6.2.1 From Classifiers to Laws, Predictions and Theories

In most sciences, results and observations do not amount to much on their own, but the goal of scientific disciplines is to establish laws of nature that explain the observations. There are many theories of science and different scientific disciplines engage in different approaches to establishing scientific knowledge, but a general framework followed by most scientific investigations could be stated as follows. Nature is observed and on the basis of these observations, tentative explanations of the underlying mechanisms that produce the observed phenomena are suggested. From these explanations, predictions are made that may be tested on data, usually data that was not used in the formulation of the explanations. Explanations that are both verified and that resist falsification become established laws and can be used to build more elaborate systems of laws called theories. From theories more elaborate predictions may be made and tested. Using similar methods, sciences can refine their theories, allowing for a more nuanced understanding of nature.

When a law or theory that is established to hold in general is falsified on a specific case, then the reason for the specific exception to the rule needs to be investigated using the same principles described above, and the theory should be adapted. Thus, scientific theories are never immutable and monolithic, but are dynamic in that they are constantly tested and revised.

The discipline of art history is quite obviously not a science in this sense, and neither is attribution study, even though it may use scientific methods. Even supposedly scientific methods like that of Morelli are not so much scientific as empirical. They were neither established by the methods described above nor are they continually modified. Instead, art historical theories of attribution (the few times they are explicitly state) are generally stated with some form of motivation, but are seldom tested. More importantly, these theories are generally monolithic and are not changed in the face of new information. Thus they lack the dynamic aspect of scientific theories.

However, as has been explained in the literature review (1.4), there is a small but growing body of literature on pattern recognition in attribution, but most of these studies (the present study included) are investigations into a very narrow aspect of attribution. Often they are presented as engineering solutions to specific problems rather than as scientific studies in an interdisciplinary field. While these studies are very useful in that they establish the basic laws, they should be seen as the building blocks of a theory of attribution rather than as end products in themselves. Very few of these studies constructively build on already existing studies, but rather present themselves as superior alternatives. At least as important is that none of these studies attempt to replicate previous studies. This may be due in part to the tendency of these studies to be published for a pattern recognition audience who are most interested in new algorithms and new applications of algorithms. Consequently there is pressure on publications to offer novel machine learning algorithms rather than verify or falsify previous claims. Furthermore, in such journals, the aim is not to further the scientific discipline to which the algorithms have been applied, but to further the science of the algorithms themselves. The remedy to this is to create specialist journals where motivation for specific algorithms is derived not from the mathematical elegance of the learning algorithms, but from art historical theory and from the findings of previous literature.

6.2.1.1 Establishing Laws: Verification and Repeatability

A key issue in most physical sciences is the requirement of replicating the findings of any study. Having experiments repeated in different laboratories doesn't simply guarantee against the publication of fraudulent results but both ensures that positive results are not simply the result of experimental design flaws, and that results are not simply due to coincidence (which can be expected to happen quite frequently when the typical α-significance level is 0.05). The same should be advocated in the study of attribution. A particular concern here is that sample sizes for most experiments in attribution will be quite small. This means that performance estimates have very high variance and therefore require independent verification. Furthermore, follow-up experiments can also be used to refine theories on which the previous experiments are based. A particularly important issue here is establish the range of circumstances under which the laws proposed in the original study hold. The study of attribution should, therefore, follow the practice of other sciences in regard of the skepticism toward claimed reports in the literature and more attempts should be made to replicate the results in any study.

As has already been stated, the aim of scientific inquiry is to explain phenomena rather than simply to describe them. Doing so requires theories to be built around well-established laws. Laws in attribution studies may be established in a number of different ways. The most obvious is when a classifier is designed that is capable of distinguishing painter X from Y. If no art historical motivation for the success of the classifier is provided (it may have been entirely motivated by machine learning principles and concerns) then the most that can be claimed by the success of the classifier is that "classifier A may be used to distinguish between painters X and Y". We will refer to laws of this kind as **trivial laws**. While these laws are often useful in themselves, they should also be seen as the starting point for further research. More useful laws are those in which motivation is provided from art historical theory or even artistic intuition, and not only from machine learning or pattern recognition principles. In such cases, not only is the law itself tested by the empirical evidence, but the underlying motivations are also implicitly verified. However, caution is required because the motivations could be post-hoc justifications for a method the researcher believed would work for other reasons. Experts are notoriously poor at expressing the rationale behind their judgments and for this reason it is also essential that these laws undergo independent verification.

Verification in this sense can take two forms. First, the trivial law may be verified by repeating the experimental procedure described in the original publication, but using different data. This should not only take the form of the data from the same agents chosen by the original authors, to ensure that the results were not due to artefacts in the experimental setup. But the experiment should also be repeated using agents that were not used by the original authors. In this case, the new agents should be those for whom the classifier should be expected work, given the motivation behind the original classifier design. In this case, it is the scope of the law's applicability that is being tested - does it hold in general or simply in the specific case of the agents on which it was originally tested? From these laws, more complex theories may be formed. Laws and theories will usually be more easily stated in statistical terms rather than absolute terms such as in physics, and they may be limited in scope in many ways.

Examples from this dissertation will now be used to illustrate these concepts. First, a trivial law derived from this study is that the classifier designed in chapter 3 is capable of distinguishing between the members of three form sets associated with Exekias, Group E and the Princeton Painter. This law was not designed only to distinguish between these specific form-sets, as is obvious by the fact that they were selected by a pilot study conducted on a different group of form-sets. Therefore the law implicitly claims a general applicability. The law may be verified in subsequent studies, by testing the same classifiers on knees by different agents, and also on different types of forms, such as ears, mouths, etc. Such a study would not only verify this law, however, but also determine its scope (i.e what types of form-sets is it good at distinguishing between).

This example is to some degree quite contrived since trivial laws are only of limited interest. Instead, the underlying principles are of more interest. In the previous example one may determine whether Morelli's principle (the art historical principle underlying the law) holds for other agents than those selected in the original study. The scope need not be limited to the medium of the original study, so for example, determining if and how Morelli's technique could be applied to Renaissance art could be of interest. Another example of assessing the scope of a principle could be derived from the findings of chapter 4 in which it was suggested that relative proportions and angles between cranial features of male subjects in Attic black-figure may be used to map stylistic influences between agents. Since classifiers trained on an analogous feature space appeared to distinguish between heads rendered by different Japanese print artists, it is worth finding out whether the same features can be used to map stylistic influences within the Japanese Ukiyo-e tradition.

6.2.1.2 Dealing with small samples

A second reason for replicating experiments is to deal with the variability inherent in small sample studies. Future studies are likely to face the same difficulties regarding availability of data, even if the measures suggested in the second part of this chapter are widely adopted. This means that the variance of the estimates for the classifier performance will be high. This not only means that point estimates for performance may often be inflated (and equally often overly pessimistic), but it also makes it easy for experimenters to present results that have been fine-tuned to the data they have collected. For example, an experimenter who tries 20 different techniques, only one of which appeared to be significant at the 0.05 level (to employ traditional hypothesis testing terminology), odds are that this positive result was simply the result of chance. Replication of experiments provides some insurance against this practice both because results will need to be confirmed before being accepted and because replication increases the likelihood that blatantly fraudulent practices will be exposed. In addition, replication also protects against results that have been skewed by artifacts in the experimental design.

A second precaution that must be taken when using such small samples is that point estimates of classifier performance should not be reported alone, but some description of the distribution of the estimates needs to be reported as well, such as the variance. The combination of replication of experiments as well as the reporting of the variance of the estimates will also allow scholars gain some idea of the stability of classifiers themselves (something that cannot be done in a single small-sample study). This can be useful for at least two reasons. First, forensic art historians may be more interested in the results of classifiers that are stable because they can reliably report their confidence in an attribution. On the other hand, in the construction of ensemble methods, unstable classifiers with high accuracy may be more desirable than less variable classifiers, since their errors are more likely to be diverse.

6.2.1.3 Negative Evidence

Laws may be established through both positive and negative evidence, so it is important that negative results should also be published. Typically this would mean that failure to replicate an experiment is of some importance in the establishment of a science of attribution and should therefore be worthy of publication. This is particularly important in cases where there is good motivation that the experiment should not fail, and where the failure in a specific case may be an opportunity to refine the theory. Negative evidence may also be used to make sense of failures and successes within a single experiment. For example, chapter 4 provides an example of negative evidence revealing an interesting phenomenon: that the confusion matrix of the classifier designed to distinguish between multiple agents based on analysis of the relative proportions of the facial features, may provide a measure of how closely the agents are related to each other stylistically.

6.2.1.4 Modeling the attribution process

The no free lunch theorem, introduced in 2.4.1 suggests that learning may only occur when there is an inductive bias, and serves as a reminder of the importance of having a model before learning can take place. When applying pattern recognition to physical problems, the laws of physical science may be used as the basis of these models. However, this prior scientific knowledge is only of limited use in the context of this study. Instead, the artistic process was largely modeled statistically by making a number of assumptions based on art historical intuition. However, art historical theory and intuition are not the only means by which particular models may be derived. It is possible that further study into the psychology, cognitive science and neuroscience of the artistic process may reveal the underlying mechanisms that govern stylistic choices and behaviours of artists, and that knowledge of these areas may be incorporated into more complex classifier designs (for example hierarchical methods like classification trees and graphical models).

There has already been some research conducted into artistic style in neuroscience, and some of it by very important scholars. For example, Ramachandran and Hirstein [1999] have argued that neuroscientific principles may explain stylistic features in certain art historical traditions, and more importantly, can explain some aspects in the development of art. For example, they explain the tendency to exaggerate the "essence" of objects (i.e. what defines them as distinct from other similar objects) by the peak shift effect [Hansen, 1959]. This phenomenon is exhibited by learners who, given two intradimensional stimuli S- and S+ (i.e. stimuli of the same kind but which differ along some dimension) and rewarded only when they respond to S+ will not only learn to respond to S+, but will respond even more to a novel stimulus S++ that differs from S- more greatly than S+ (but in the same direction as S+). Ramachandran and Hirstein [1999] give the example of rats rewarded if they respond to rectangles (S+) but not to squares (S-) will respond more strongly to novel rectangles (S++) in which the ratio of length to breadth is more than that of the rectangles to which they've been trained to respond. They see the exaggeration of the female form, for example in a Chola bronze statue of Parvati,[2] as an instance of artistic style expressing the peak shift effect.

Ramachandran and Hirstein [1999] have been criticised for a number of reasons, most notably because evidence for a particular cognitive effect in artistic style may be offset by the appearance of the opposite tendency, sometimes in reaction to it [Tyler, 1999] and because the peak shift affect only works in rats when the difference between S+ and S- is small [Martindale,

[2]Saraswati Mahal Art Gallery. Photograph from Behl [2008].

Figure 6.1: A bronze statue of the Hindu goddess Parvati with exaggerated bust and hips, and narrow waist, believed by Ramachandran and Hirstein [1999] to be an example of the peak shift effect expressed in art.

1999]. An example that illustrates how difficult it is to test Ramachandran's claims is that the exaggeration of female proportions visible in the Chola bronze is matched by frequent androgyny and hermaphroditism in the same tradition. It is thus very difficult to prove that the exaggeration of the female form is due to the peak shift effect when it is not universal. Of course, the usefulness of neuroscientific approaches is that there is always the possibility that neuroimaging of artists may provide some way of falsifying these theories.

Ramachandran's theory is only one example of attempts to model art from the perspective of the cognitive disciplines. Such attempts have existed since at least Hambidge [1920]. Much of these studies have focused either on prescriptions for 'good' art or on finding a handful of features that explain the commonalities of all art. Although a unified theory of art is desirable it should account for, rather than ignore, the differences between artists and movements. Nevertheless, by making the assumption that these neural or cognitive mechanisms that underlie the creative urge, are expressed slightly differently from artist to artist then at the very least, these neuroscientific principles may tell us where to look. For example, in areas in which we should see the peak shift effect occur (such as areas in which a particular tradition regularly exaggerates a particular form) then a study of the extent to which these forms are exaggerated may reveal diagnostic criteria for attribution and investigating whether this is the case may be fruitful areas for future research.

In some cases, research into automatic attribution can be based on empirical art historical research (as opposed to art historical theory). I will use one example to both state and exemplify the case. Mary B. Moore conducted doctoral research in the anatomy of horses in Attic black-figure in which she established a set of categories associated with each anatomical detail, and documented the painters that were associated with respective categories [Moore, 1971]. This will be illustrated with one example (from many in the work) the eyes. Moore distinguishes between a number of different types of eye. For each she lists the paintings in which these types are expressed and in a table that lists the painters who frequently use them. The example in figure 6.2 shows two examples, the type V and and type XI.c - the former only associated with Group E during the 3rd quarter of the 6th century, and the latter only associated with the Princeton Painter during the same period. Using the terminology developed in chapter 2 one may describe the categories of tear duct as form-sets and each instance of a tear duct as a form. Moore's dissertation covers many other anatomical details in the same way, including manes, shoulders, chests and many others. Because she lists all the vases from the set she sampled (1000 vases in total) in which each feature

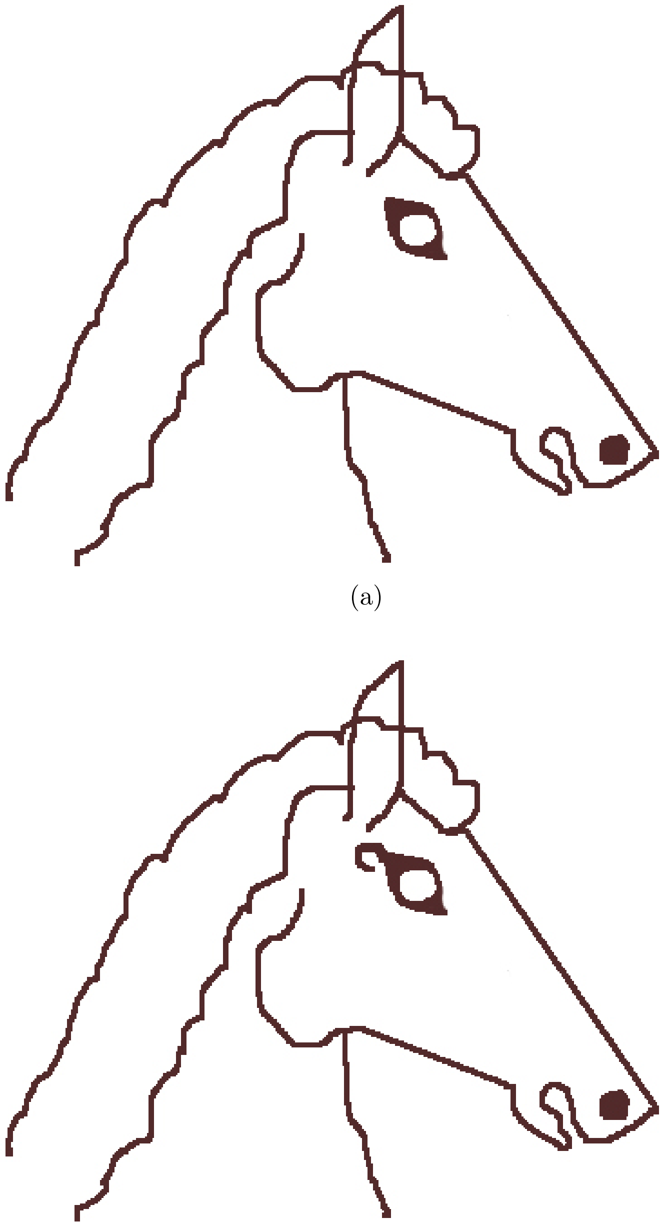


Figure 6.2: Illustrations of two types of eye that appear on Attic black-figure horses. (a) is associated with Group E and (b) with the Princeton Painter.

appears, this study may be used as a source of likelihoods of particular forms given particular painters and an example of future research building on this art historical PhD is a Bayesian analysis of horses' anatomical details.

6.2.2 Combining Approaches

The preceding section described how a science of attribution could develop, using the construction of classifiers as rudimentary tools for validation and falsification of laws of attribution. The consequence would be the development of an increasing number of different methods for attribution, and a proliferation of different classifiers designed for this purpose. The present section offers some suggestions as to how a machine that automates the classification process may combine these different approaches. In addition, the issue of the role of the human expert in the process is assessed.

All the examples are based on the following scenario. A form γ, possible a sherd or a vase painting, requires attribution, and that by some method, possibly by the eye of a human expert, the number of possible agents that could be responsible for the form is reduced to some manageable number, N. These agents will be referred to as the candidate agents. With each candidate agent is associated a state of nature ω which is the event that this agent is responsible for γ. The set of N possible states of nature is indicated by $\Omega = \{\omega_1..\omega_N\}$ where ω_n is the outcome that agent n is responsible for the form. A set of experiments are conducted based on classifiers that are capable of distinguishing with above chance accuracy, between the work of the candidate agents. It is assumed that each of these classifiers will rely on different features and these may be based on sub-forms of the original form γ. For example if γ is a sherd that preserves a human male figure, then two sub-forms could be his knee (chapter 3) and the relative proportions of his cranial features (chapter 4), to use examples from this dissertation.

Candidate agents may be selected in a number of ways. Fabric, medium, shape and age may be used to narrow down the possible agents. Such criteria are easy to evaluate, even with relatively untrained eyes, and it is a simple matter for archaeologists who are not familiar with painter attribution to narrow down the possible agents using these criteria. Analogously, in art from the Renaissance and later, it is quite a simple matter for an enthusiast to narrow down the possible agents for an attribution based on knowledge of the school from which the painting came (it is not likely, for example, that one would confuse Picasso and David). This knowledge would allow the enthusiast to select the appropriate set of tools to refine the attribution further.

Three scenarios are discussed in turn. The first considers the scenario in which all of the tests conducted on the form use Bayesian classifiers. The second in which the classifiers are not necessarily Bayes classifiers but, in addition to making an assignment, also provide estimates of the respective probabilities that each agent produced the form. The final scenario makes no assumptions about the type of classifiers used in the tests except that they make an assignment.

6.2.2.1 Case 1: Bayesian Classifiers

If Bayesian classifiers, such as LDA, QDA or the naïve Bayes classifier, are used as the basis for classification, then they may be used in series using Bayes theorem to update probabilities as new evidence is found. The method is explained below, first informally, and then formally using an artificial example from the attribution of Attic black figure. To recap, Bayesian classifiers make an assumption about the function that generated the data presented to it. Given some prior probability distribution over the candidate agents, the Bayesian classifier will produce, for each of the candidate agents, a posterior probability that the agent was responsible for the data, and attribution is simply to the agent for whom this probability is the greatest. If a new test is conducted and new data generated from the same form using a different feature extractor, then the posterior probability from the previous test may be used as the prior for the current. This procedure may be repeated for as many different tests are conducted.

A simple and intuitive approach to explaining these methods may be to view Bayes theorem as

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{normalising factor}} \tag{6.1}$$

The posterior directly tells us the relative probabilities that the form is the work of each respective agent, and we would typically attribute to the agent for whom this posterior is the highest. The prior expresses the relative probabilities before conducting the test, for example the art historian's own intuitive opinion. The normalising factor is simply to ensure that the sum of the all the posteriors is 1 and may be ignored from a conceptual point of view. Finally, the likelihood tells everything about the relative probabilities of each painter given only the distributional assumptions and the data itself. Thus Bayes theorem may be used when multiple experiments are conducted in series by using the posterior of the previous experiment as the prior of the next.

More formally, Bayes' theorem was stated in chapter two and this is repeated here. Assume a form γ requires attribution. Let $C_1..C_N$ be a set of $N \in \mathbb{N}^+$ classifiers operating respectively on feature vectors $\mathbf{x}_1..\mathbf{x}_N$ of γ. Let the common output space of the classifiers be $Y = [0,1] \times \Omega$ where $\Omega = \omega_1..\omega_G$ represents the G possible states of nature and [0,1] is the the range of the posterior probability that each of the respective states of nature is true. Typically, the states of nature would be the events that each painter was responsible for the form. Then by Bayes theorem the output of classifier $n = 1 \ldots N$ may be written as:

$$Pr_n(\omega_g) = \frac{Pr(\mathbf{x}_n|\omega_g)Pr(\omega_g)}{\sum_{g=1}^{G} Pr(\mathbf{x}_n|\omega_g)Pr(\omega_g)} \tag{6.2}$$

Where $Pr_n(\omega_g)$ is the probability that the form γ was produced by the agent described by the state of nature ω_g.

If the classifiers are applied to the form in series then the following procedure may take place. Equation (6.2) is used with a Prior $Pr(\omega_g) = 1/G$ for the C_1 (i.e. for the case where $n = 1$ only). Then the posterior for this value is used as the prior for the next classifier. Thus equation (6.2) for each subsequent classifiers $n = 2 \ldots N$ becomes

$$Pr_n(\omega_g) = \frac{Pr(\mathbf{x}_n|\omega_g)Pr_{n-1}(\omega_g)}{\sum_{g=1}^{G} Pr(\mathbf{x}_n|\omega_g)Pr(\omega_g)} \tag{6.3}$$

In other words, the posterior for the previous classifier is used as the prior for the next.

For example, say a vase of unknown authorship is discovered and an archaeologists narrows the possible attributions to the Painter of Berlin 1686, Exekias, The Swinger, The Princeton Painter and Group E. Then γ is the form to be attributed and $\Omega = \{$ Painter of Berlin 1686, Exekias, The Swinger, The Princeton Painter, Group E $\}$. If there exist three possible tests for attribution based on three different sub-forms, say the eyes, ears and knees, then these may be represented by the respective feature vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$. Assume that LDA is undertaken and the idea is to attribute to the painter with the greatest likelihood, then LDA would be conducted as normal using $\mathbf{x}_1$ as the data and the posteriors, $p_1(\omega|\mathbf{x}_1)$ for each of the agents is calculated. Then LDA would be conducted using $\mathbf{x}_2$ as the data, but with the following modification, the prior probabilities for each of the agents ω would be the posteriors from applying LDA to the data $\mathbf{x}_1$. The same is done when LDA is carried out on $\mathbf{x}_3$ where the posteriors for LDA calculated from $\mathbf{x}_2$ are used as the priors. Finally, attribution can be made to the agent for whom the final posterior,

$$p_3(\omega|\mathbf{x}_3) = \frac{p(\mathbf{x}_3|\omega)p_2(\omega)}{p(\mathbf{x}_3)} \tag{6.4}$$

is the greatest. In many circumstances, the results rapidly converge in favour of one of the candidate agents.

Despite having the veneer of mathematical formalism, this method is an approximation and is only coherent with probability theory if certain conditions are met. The most important of these is that the different tests be independent of each other. Intuitively this can be understood quite simply as stating that the prior distribution must actually be prior - i.e.it must in no way be conditional on a knowledge of the data at hand. Of course this will probably not be the case in real life. However, there may well be some merit in future research into the conditional dependence of various statistical tests of attribution. A second condition is that the distributional assumptions of the likelihoods must be true. Again, this is highly unlikely to be the case in real life. For example, LDA and QDA assume that the data are normally distributed and LDA makes the extremely unlikely assumption that the respective covariances of the different classes is the same. This does not mean that the method cannot be used - pattern recognition always makes assumptions about the world and approximate methods are abundant. However, this warning is provided because its formal appearance may make Bayes theorem look more attractive than it should be. More importantly, these methods are only really appropriate if the classifiers are based on Bayes theorem. In many cases, this may not be appropriate.

6.2.2.2 Case 2: Posteriors Known

Bayesian classifiers are not the only classifiers that output posterior probabilities. For example, there are methods that directly attempt to calculate the class conditional densities. One of these is k-nearest neighbour (k-NN) which has been used in Ch3 and Ch4 and which is explained in detail in 2.4.4. If a set of attribution tests are conducted on the same form, each producing a different posterior probability, then these may be combined by simply summing the posterior probabilities for each of the candidate agents and simply attributing to the agent whose sum of posterior probabilities is the greatest. Should it be desirable to know the probability that a particular agent is responsible for the form, this can be achieved by dividing by the sum of the posterior probabilities for each of the candidate agents.

6.2.2.3 Case 3: Posteriors Unknown

There are many approaches to classification, however, that do not attempt to provide probabilities for each of the states of nature but simply selects one candidate agent over the others such as, for example, the ensemble methods used in this study. However, in such cases, ensemble methods, similar to those used in this dissertation may be used to combine the outputs of different classifiers. A key observation here is that if the classifiers are designed to look at different aspects of a vase, they are very likely to make independent errors from each other. Independence is a key factor in the success of ensemble methods, and indeed, we would recommend that in such cases, the results of classifiers should simply be combined, either by majority vote, or by more complex ensemble methods. There are some weaknesses in using majority vote ensembles in cases in which the number of component classifiers is small in relation to the number of classes. One such problem is that there may be ties for class membership, in which case a principled rule is required to break these ties. One possible area of future research is into ensemble methods for design of classifiers based on different base classifiers. Such methods need not be as simple as majority vote methods, but may be hierarchical incorporating a number of art historical principles. This point is discussed in more detail in section 6.2.2.7

6.2.2.4 The Role of Expert Knowledge

At some point in the future, computers will be able to engage in practices like attribution with greater accuracy than a human but it is impossible to predict when. A cursory examination of the literature on computer-based attribution of art suggests that it won't be in the near future. One may argue that scientific authentication, such as is sometimes used in legal cases in which fingerprints on the frames[King, 2000] or the chemical composition of the media are analysed, amounts to an automatic method of attribution that surpasses a human's ability. But that would miss the point that so called scientific attribution is conducted by a human who selects ad-hoc scientific methods in order to conduct the attribution. In fact, scientific attribution is a concert of human and machine, not a fully automated method. A fully automated method would not require a human user to select the appropriate features or scientific methods. This is still a long way off.

This is not to say though, that the methods and techniques presented in the literature on machine based analysis of paintings, and indeed the methods presented in the present study, can't be applied today. However, the question should be asked: would we always trust the machine to make the correct decision? If not, what role should experts play in the use of such algorithms on real examples. A complete answer to the question is beyond the scope of this study, but some salient issues will be discussed here. First in a general and theoretical way, and then practical guidelines for the implementation of methods such as those presented in this study are suggested. It is acknowledged that these are properly topics for future research, and the most that can be achieved in this chapter are suggested ways forward.

6.2.2.5 Implicit Expert Knowledge

It is easy to overlook the degree to which expert knowledge is already encoded in the methods that have been proposed in the literature, and in the present study. First, and most obviously, the process of feature selection and extraction is strongly guided by human experts. We should, for example, be very surprised if a classifier would be able to make sense of a painting if the feature space consisted simply of the individual RGB (red-green-blue) pixel intensities of a digitised image of the painting. Instead, the algorithm would have to be pointed in the direction of salient features to extract from this image for anything meaningful to be induced. This is illustrated in the literature, for example, by models that work on brush-stroke analysis [Sablatnig et al., 1998, Melzer et al., 1998, Yelizaveta et al., 2006a,b]. They are predicated on the belief that different painters develop different techniques for applying paint to the canvas and that any technique that is associated with a specific painter is also applied with some consistency by that painter. This assumption is not justified by first principles nor fully articulated within art historical theory, but it has been common knowledge and general practice among art historians that close examination of brush strokes can often reveal the identity of different artists. Thus, some feature space defined on the brush strokes is the basis of the attribution, and this choice is based on human expert knowledge.

A second area in which human expert knowledge is implicitly incorporated into the classifier design is the construction of the set of class labels. This has two implications. First, the set of candidate agents to which attribution can be made, is usually chosen by the human expert. All of the methods prescribed by the literature limit the scope of the attribution dramatically. Indeed, a machine that can correctly attribute new works of art by itself to any possible painter would be required to know the identities of hundreds of thousands of artists and a 100 000 way classification procedure is difficult at best, but much more so when the sample is so small - many of these artists have extant outputs in the order of 10 works. Clearly, when a machine is used for attribution, the number of possible predictions has to be limited in some meaningful way, and this limitation should be imposed, at least for the moment, by the expert. Secondly, it is the human expert who has to label the data in the design set. In some disciplines this may be as simple as examining the signatures on the works or consulting the relevant documentary evidence.

However, there are still large corpora of unattributed paintings (or we would not be engaging in this exercise) and labeling the design set in this case has to be done manually. In this case, the training phase will bias the learning algorithm which will implicitly include the knowledge and intuition of the expert that labeled them (Beazley in the case of this dissertation).

6.2.2.6 Expert Probabilities

A direct way in which to incorporate expert knowledge is to elicit from them their degrees of faith in the different possible candidates. This may be as simple as stating their preferred candidate or as complex as specifying a distribution over the probabilities of certain candidates being responsible for the form. If a prior distribution of some form is specified, then this may be incorporated as a prior probability in a Bayesian classifier like LDA. One of the major problems with this method is that it is difficult to elicit priors from an expert, particularly one with limited knowledge of statistics. A number of methods have been used to elicit both point priors (i.e.single values describing prior belief in each agent) and continuous distributions, a summary of which may be found in Jenkinson [2005] These include using methods such as wagers to evaluate the experts true opinion of the respective likelihoods. This is a rich field, the literature for which is growing. However, it is worth pointing out that even unreasonable information encoded in a prior will often be overwhelmed by the data after only a few experiments have been run.

6.2.2.7 Ensembles

Another possibility is for human decisions to be used as part of an ensemble of classifiers. This may be done in a number of ways, the simplest of which is simply to treat the human as one of the base classifiers in some kind of majority vote ensemble - perhaps giving the human expert a higher weighted vote than any of the other base classifiers. However, while this dissertation has used majority vote ensembles, there are more complex methods of combining base classifiers and it may well worth studying how human experts may be used to create such ensembles. A very active area of research in machine learning is in solutions that impose structure on the problem space in various ways, such as decision trees. These methods can often reveal considerably more about the classification process than more opaque methods like LDA (which is analytically very simple, but does not reveal much structure in the data) or multi-layer perceptron neural networks (which are analytically hard or intractable). In particular, decision trees may be used with multiple classifiers that are specifically trained to handle very narrow specific tasks. There are many different types of decision trees including some methods for which the specific algorithms are not public domain (such as C4.5 which is used extensively by linguists and researchers in natural language processing). The idea behind most of these is to split the classification task into multiple sub-classification tasks that are arranged in a hierarchical structure. The tree structure may be described by a set of nodes, a set of leaves and a set of branches. A node represents a classification task, the outcomes of which are represented by branches. Each branch may lead to another node and more branches. A branch that does not end in another node terminates in a leaf, which is a prediction. While the construction of a classification tree may be conducted entirely automatically by a supervised learning process, expert knowledge may be used to simplify the learning process, such as in serial combination of multiple experts.

Research of this kind is already being conducted elsewhere and a good example will be used here as an example of human expert knowledge encoded into a hierarchical ensemble classifier. Yelizaveta et al. [2006b,a] use a hierarchical classifier as an intermediate step in the process of high level semantic labeling of a painting (attribution is an example of high-level labeling). They describe a system that assigns labels to the brushwork according to which of 8 painting techniques it belongs, such as mezzapasta (a painting technique in which paint is applied with medium thickness) and pointillism (the paint is applied with many small dots, rather than with brush strokes). Different feature spaces are associated with distinguishing different methods from each other and on this basis, a hierarchical combination of expert classifiers was designed to recognise each of these. At each node, a different classifier was used to refine the classification, for example, to distinguish between Impasto and Divisionism, colour contrasts and wavelet pyramid transformation was used. Figure 6.2.2.7 is a graphical representation modified from Yelizaveta et al. [2006b] that illustrates the process. This kind of hierarchical ensemble implicitly encodes human expert knowledge since it is humans that structure the solution space (traditional classification trees do this automatically). Such approaches may well have some application to Greek art, and there is considerable scope for investigating this issue.

6.2.3 Summary

The first section of this chapter has looked forward toward future research into automated and semi-automated attribution, not only in Greek art, but in art history generally. The main suggestions are that future research be conducted using a scientific method. In this regard, the ultimate aim of

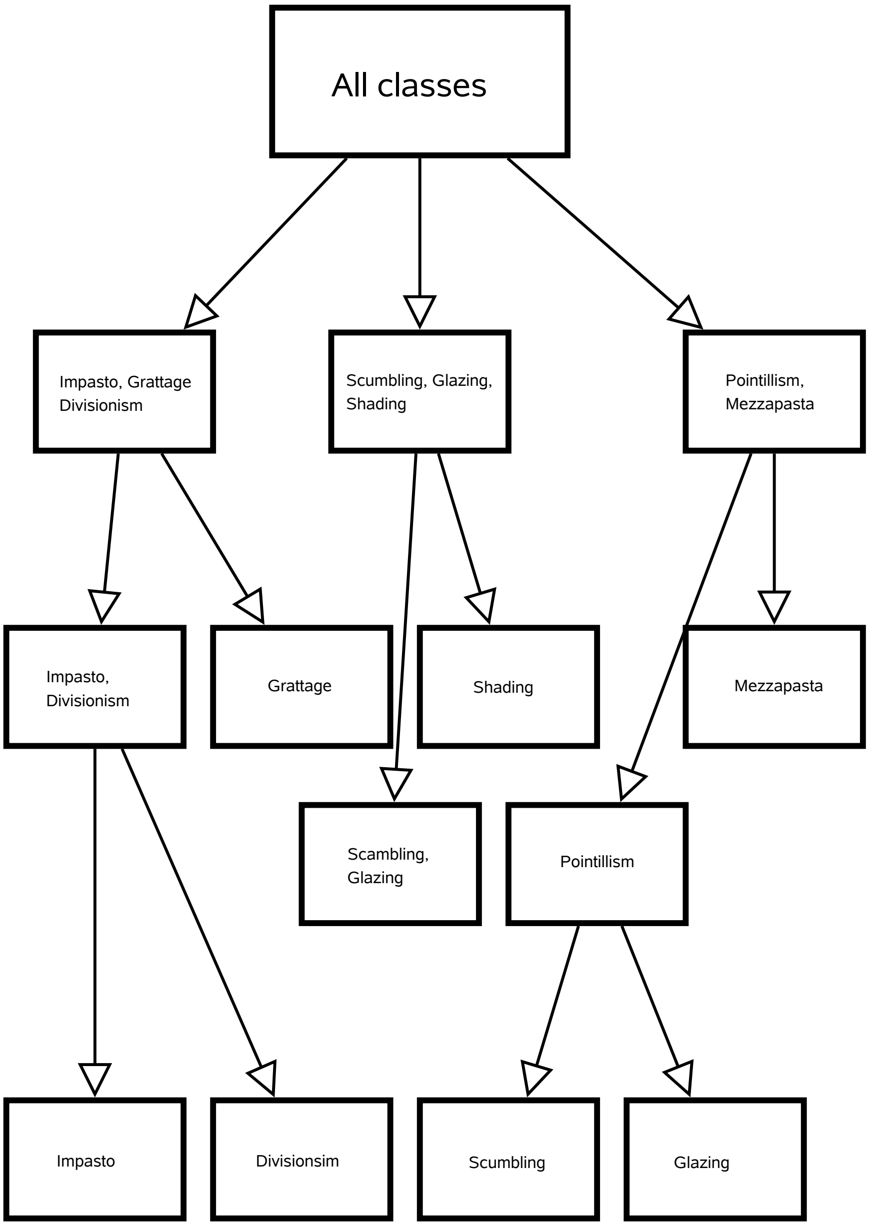


Figure 6.3: A diagram illustrating a hierarchical classifier that trains specific experts to distinguish between a narrow set of features and combines these in hierarchical ensemble. Modified from Yelizaveta et al. [2006b].

this scientific pursuit is the unification of attribution in different media. In addition, particularly close attention should be paid to advances in cognate and relevant fields, such as art history, cognitive science and neuroscience. In particular, these fields are beginning to research the artistic drive and they may help in the establishment of models on which to base attribution. The result of such an endeavour will be the development of a number of different classifier designs and methods which leads to the question of how these different approaches should be combined to produce an attribution. A related issue is what role the human expert should play, if any, in the attribution process. A number of suggestions include through the elicitation of expert probabilities and treating experts as base classifiers in ensembles. However, the biggest role for the expert is likely to be in the process of designing the classifiers themselves.

6.3 Dissemination and Publication of Attic vase-paintings

The first section of this chapter made suggestions for further research into classifier design. This second section deals with addressing a key issue that impacted negatively on this study: the availability of good quality data. In particular, it suggests certain changes to the way in which Attic painted ceramics are published that will not only facilitate further studies similar to this one, but will be of general benefit to the scholars conducting more traditional studies. In particular, the paucity of good data and images for the analysis of vase-painting placed considerable constraints on the way this present study was conducted. This study is intended as a proof of concept rather than a fully implemented system, but the implementation of a fully functioning digital attribution device would require data of considerably higher quality than is available. It is argued here that the *Corpus Vasorum Antiquorum* (*CVA*), while a very useful tool for scholars, may be improved by taking advantage of the availability of new technologies. This section is intended as a critique of the current situation and an appraisal of various proposed and emerging solutions to the problems. Although the implementation of such a project is beyond the scope of this dissertation, a number of recommendations are presented here together with a motivation for this change and a discussion of the issues that need to be considered when decisions need to be made about how to implement improvements or alternatives to the *CVA*. The rest of the section will proceed as follows: after a brief motivation (6.3.1), a brief critical survey of current practices of electronic disseminating of data (6.3.2), and a discussion of three dimensional techniques that may be used for future dissemination of data (6.3.3).

Attic ceramics are published or disseminated in a variety of ways. First, in archaeological reports where information is recorded for the pieces and drawings and profiles are taken. This will also include some specialised information about the artefacts which ranges from vague discussions to very detailed analyses. Secondly, these pieces may find their way into the market and in the most fortunate of cases, into museums. Here they may be published in the *Corpus Vasorum Antiquorum* project where respective museums publish their entire catalogue of Greek ceramics more or less according to an established format that includes black-and-white photographic plates, profile drawings, and descriptions that (usually, but not necessarily) include the shape and type, the condition, the dimensions, a brief description and interpretation of the scenes if any, the subsidiary decoration, an attribution (if there is one), a bibliography (if warranted), and the added colours (i.e.red and white) and inscriptions. By the time this occurs, many of the pieces may already have been published by the art dealers in various catalogues (although the information presented there will usually not include anything that was not in the archaeological report). Many museums also have in-house journals or published catalogues in which certain pieces will be published, sometimes with more substantial discussion of the scenes than in the *CVA*. Finally, the pieces may also be published in scholarly journals either as the exclusive subject of the paper or as pieces of evidence for a larger issue relating to iconography or style, and as evidence in more general Classical studies. The main problem is that the standard used is still black-and white photographs, the quality of which is highly variable, as are the conditions under which these photographs are taken.

6.3.1 Motivation

There are a number of reasons that the manner in which Attic ceramics are published should be revised. Some of these are motivated by the idiosyncratic requirements of studies such as this one but which, if addressed, will be of general benefit for future studies involving machine learning and pattern recognition applied to this discipline. Others are motivated by a more general lack of access to, and quality of, available materials, of which the former complaints are merely symptomatic. Finally, this section is motivated by the fact that while technologies exist to ameliorate both types of problem, these are not being adopted in a widespread manner for reasons which have little to do with the cost or complexities of these technologies, but which are related instead to structural and logistical issues in the manner in which

Figure 6.4: A photograph of a scene from Basel BS 427 **EM1**. The quality is very high but it is difficult to recognise the Morellian features of the figures near the edge.

museums are run, scholarship is undertaken, research funding is allocated and information is disseminated.

The techniques presented in this dissertation suffered from severe poverty of data, which was neither good nor abundant. In this dissertation, the strategy adopted to ameliorate the problem was to sacrifice flexible models in favour of those with low variance. An increase in the amount of data available would have allowed more flexible models to be used and possibly increased the performance. Furthermore, with better data, a wider range of techniques may have been used. Two examples of areas of inquiry which have proven useful in the attribution of art in other areas, but which were not possible in this study because of a lack of adequate data, should serve to illustrate the point. First, the photographs available did not preserve textural information which may have been available if higher resolution images had been available. Textural information has been useful in other areas of art attribution (such as Li and Wang [2004],Jiang and Huang [2004] and Melzer et al. [1998])[3] Typically this was to analyse brush-strokes, and while brush strokes are not typically visible on Attic vases, there are other textural features that could be of some interest. One such possibility is an analogous examination of

[3]These are discussed in the Literature Review in 1.4.3.

the texture of incisions, which was impossible given the scarcity of close-up photographs of the incisions. Secondly, it is clear that some painters have a distinctive approach to composition, such as Exekias' knack for elegant simplicity of composition, the Amasis' painters love of strong vertical and horizontal emphases, and the Affecter's preference for monumental figures. In other fields of art attribution and analysis, formal and computer-aided studies of painting composition have been undertaken (such as Tanaka et al. [1999]). However, because of the curvature of the vases and the fact that photographs of the scenes were not taken from a consistent angle, it was difficult to undertake a meaningful study of composition. The main difficulty is that on typical images of a scene (such as in figure 6.4),[4] the edges of the scene are distorted to the point that even with very high resolution images, there is little informative pixel data at the edges. Typically (i.e.in standard non-computer aided analyses), these problems would have been overcome by *ex vitrine* examination of the respective vases by the respective researchers. However, with the amount of data required for supervised learning, this would be prohibitively expensive for most scholars.[5]

Although the arguments above have been illustrated by issues that were of concern only to this project, they can be generalised to the study of artefacts using machine learning techniques in general. To this effect, similar arguments were made by Durham [1996] over ten years ago. Of more concern than a lack of data available for the present study, the aim of which is in any event simply proof of concept, is the impact on future studies, the data requirements for which are not be easy to predict. Thus, it is important that future standards for publication of Attic ceramics should contain as much visual information as funds and current technology allow. What is certain is that if the number of studies in computer aided art analysis increases, so will the demand for good quality data. More importantly, machine learning and pattern recognition applications are far from the only areas of inquiry that would benefit from using up-to-date technology for the publication and dissemination of images and data. The fact that expensive *ex vitrine* examinations of vases are still being carried out at all suggests that better quality data should be of use to a broader audience. Thus, changing the way in which artefacts are published will be of use not only to those interested in computer-based studies like the present, but also those conducting more traditional research into vase-painting.

[4] Photograph courtesy of the Antikenmuseum Basel und Sammlung Ludwig

[5] In the case of the data required for Chapter Four, for example, 50 vases were used which are housed in 20 different museums located on 3 different continents. The cost of travel to each of these sites would have been in the order of a hundred thousand US dollars.

6.3.2 Current Practice

6.3.2.1 The *Corpus Vasorum Antiquorum* Project

The Corpus Vasorum Antiquorum is the standard resource for good quality photographs of Attic vases. The corpus is published in numerous volumes under the names of the collection in which the artefacts are housed. The research, descriptions and photographs of these objects are courtesy of the respective collections, and are often based in the first instance, on the museum catalogues, supplemented by expert analysis and photography. The *CVA* now has some 300 volumes available and represents a substantial portion of the worlds largest collections. Most of these volumes are available on-line[6] and the plates and descriptions are linked to the database of the Beazley archive[7] such that scholars using Beazley's lists as the basis of their research are able to link to the relevant pages of the on-line *CVA* from Beazley archive search results. These two resources together are the lifeblood of research into iconography and style in Attic vase-paintings.

The *CVA* project had its origins in the Union Academique Internationale, a primarily archaeological society that aimed to unite scholars around the world in joint projects to the benefit of the whole world rather than in the narrow interests of any single country. Edmund Pottier, professor at the Ecole, proposed a project in which all ceramic objects would be published by the museums in which they were housed, and according to a standard prescribed by the Union. While the project's scope was soon reduced to just Greek wares, the project gained momentum and is one of the most valuable resources available today. Pottier's main reason for suggesting the project was precisely the reason provided in the motivation above - that there was insufficient access to good quality images of the vases. The solution adopted was more than sufficient for the time, given both the technology of the day and the uses to which scholars would put the *CVA*. The standard of black and white photographs together with descriptions was ample for most scholarly purposes, and of course, those requiring more would either request better quality images of the vases they were researching or would make the expensive *ex vitrine* examinations.

The author believes that the method in which the *CVA* is published, particularly now that most of the expensive volumes are available online, is highly laudable, although more fascicles should be released from copyright and the images should not be watermarked. Of more concern than the database and its accessibility is the quality of the photographs that are

[6] *CVA* Online: http://www.CVAonline.org.

[7] http://www.beazley.ox.ac.uk.

taken by the museums. In particular, many techniques exist that allow good quality three-dimensional images to be scanned and disseminated in digital format. Doing so will largely negate the need for scholars to study these objects *ex vitrine*. More importantly, if these images are made available by a central database, then large amounts of data may be downloaded easily from a single portal, making the construction of test and design sets for machine learning approaches to attribution feasible.

6.3.2.2 Online Resources

There are a number of alternative approaches to publication of ceramics that are currently employed outside of the auspices of the *CVA*. The list is large and a brief summary rather than an exhaustive survey is provided. The summary furthermore only comments on electronic repositories or protocols. The most important is the Beazley archive which is the standard online database for researchers in vase-painting. The Beazley archive allows users to conduct very narrow searches for vases according to search criteria that are meaningful to art historians. These include fabric, shape, decoration (including keywords describing the scenes), artist and scholar. The archive keeps an up-to date bibliography of images of each vase and links to the relevant *CVA* photographic plates and descriptions. In addition, the Beazley archive has its own collection of heavily watermarked images, including museum photographs and some of Beazley's own photographs and drawings, which may be accessed via the archive database. A second large database of images is hosted the Perseus project at Tufts University. This has a large database of images of vases as well as a large selection of Classical texts in original and translation.

There are many other databases that are not limited to vases, but also offer access to other kinds of archaeological data. Some excavation sites, for instance, publish three dimensional images of some of the finds on the web. Often these are in the form of object movies (**??**) that allow the user to rotate the object around a single axis. In some cases,[8] the objects are simply made available by browsing the web-page, but in others there is a complete and in some cases extensive database that may be searched.[9] A interesting innovation is the increasing use of flexible databases that allow scholars to make their own excavations available via a single portal, such as opencontext. These innovative techniques are not only limited to the artefacts but in many cases the actual excavations can be explored using panoramic photography,

[8] such as the Herculaneum Project: http://www.proxima-veritati.auckland.ac.nz/Herculaneum/objects/index.html.

[9] For example the Great Petra database: http://proteus.brown.edu/PGTdata/Home.

as is the case with the Herculaneum project.

In addition to these open innovations, which are limited to certain excavations, there are museum databases, many of which include some three dimensional images. A good example is the University of Melbourne's Ian Potter Museum that has published QTVR images on a CD included with the catalogue[Conner and Jackson, 2000]. The British Museum has an open database, Compass, which provides descriptions and images of a substantial portion of the museum's collection. Although the images are of high quality, they are mostly 2D photographs rather than 3-D images. While there are other institutions that have databases of good quality, these are typically only available to members of staff. What does hinder the publication of good quality three-dimensional objects is often not the cost or the practical difficulties, but rather intellectual property rights and copyright. A more complete discussion of the difficulties involved in obviating these difficulties is provided by Smith et al. [March 21, 2008] and there is certainly room for more scholarly activity in addressing this problem.

Apart from opencontext, the methods described above have a top-down description system. That is, the descriptions and search fields are described by experts and users have to make searches according to these criteria. Recently as web 2.0 interactive technologies have become more widespread, bottom up tagging approaches have gained popularity not only in popular media repositories like del.ic.ious and flickr, but also amongst the museological community. These approaches generally provide a portal that allows the user to search multiple databases using their own descriptions to organise their data. Typically this means that the end user may tag data and their tags may be shared with other uses. Thus, the data will typically end up being organised according to the needs of the users rather than the experts. Two examples of are the Virtual Lightbox for Museums and Archives [Smith et al., March 21, 2008, 2005] and the system used by the *Art Museum Social Tagging Project.*[10] To a large extent these systems have been aimed at satisfying the needs of general users rather than scholars and the motivation provided by their proponents are largely pedagogical rather than research oriented.

The task of designing and implementing a database that allows access to high quality images of vase-paintings that may be used for research is not a trivial task. Issues that will have to be addressed in the database design include whether top-down or bottom up approaches to tagging and indexing are used, whether there should be a single centralised repository

[10]More information can be found at the Steve Museum home page: http://www.steve.museum.

(such as the Beazley archive), whether links to individual museums' own collections are made available through a portal, or whether some of the newer technologies like the VLMA be adopted as a standard. It could be argued that folksonomies might provide too much flexibility and be detrimental to specialised studies in vase-painting. In particular, early in the development of folksonomies it became clear that a number of problems beset the approach. In particular terms were imprecise and personal to the point that there was a proliferation of synonyms and single terms dominated. However, it has recently been shown that this ceases to be the case as the number of users increases. For example, Guy and Tonkin [2006] shows that only 15% of the tags in flickr are single word tags. However, since the number of users of a database of Greek vases is likely to be small and it may be that "critical" mass is never reached, and the problems that beset bottom-up approaches to metadata may plague such systems when applied to such specialised domains. On the other hand, top-down single portal approaches lack flexibility and cannot be extended when the need for new search criteria arise. For example, the Beazley archive does not have fields for rich descriptions of the data, such as those generated by automatic classification of the images.

Despite these disadvantages, central portals like the Beazley archive have the major advantage that searches are aimed at the needs of specialists and that this makes it easy for specialists to build a representative sample for any specific study they wish to conduct. For this reason, it is likely that the best solution would be for a centralised repository like the Beazley archive to be used alongside more flexible systems for those who desire the flexibility they allow. In this regard, methods that allow bottom-up and top-down approaches to be integrated is an active area of research (for example [Voss, 2007]) and their application to the problem at hand is a rich area for future research. For the moment though, the Beazley archive's established infrastructure is well known in the community and provides the best method for the widespread dissemination of high quality 3D models of the vases. After all, a new portal would require buy-in from the community. This is all the more significant since the improvement in image quality offered by 3D systems is expected to be of benefit to the broader archaeological community not only those who are up-to date with current online trends. Since *CVA*online is already integrated with the Beazley archive's search engine, the most immediate solution is for *CVA* to encourage the use of 3D models for the publication of images and for these images to be made available via the online repository.

6.3.3 Image Acquisition

Thus far, the method by which the data is accessed has been discussed. This is, to a large extent, not the main problem since the Beazley archive and the *CVA* have made accessing the data very easy. The main concern is that the data is of poor quality, and the chief gains to be made will be in revising the methods by which the images are captured. The tradition in the study of Attic black-figure has been to use small or medium format black-and white photographs for the images of the vases. The manner in which the photographs were taken was very variable at one stage. However, since the mid-fifties there has been some standardisation of the procedures owing in part to Hansjörg Bloesch's recommendations, as discussed in 5.3.2. Photographs are typically taken from such a position as to give a good reflection of the vases' shape. In most cases, close-up photographs of the scenes are also taken, and in rarer cases, close-ups of interesting elements of the scene are also published with the vase. For the reasons outlined above, this is insufficient for the purposes of many of the techniques employed in this study. Instead, there exist many three dimensional techniques that could possibly solve many of the problems faced in this study, and indeed would be a vast improvement over the current system for all scholars of Greek pottery.

6.3.3.1 Current State-of-the-Art

Current state of the art is described by Mara et al. [2007] who were commissioned to develop a state of the art system for digital documentation of Attic polychrome vases of the sort treated by the present dissertation (although red-figure and white-ground as opposed to our black-figure). The method used was to capture the vase using a Konica-Minolta structured light scanner to obtain the 3D objects and high-resolution surface texture. In addition, multispectral analysis was conducted in order to determine chromatic characteristics that are invisible to the human eye. These were then registered onto the 3D model for a composite and complete description of the objects. The applicability of the techniques will be illustrated in a forthcoming volume of the *CVA* in which the profile drawings and 2D images are rendered from the 3D models.

There is little disputing that Mara et al. [2007]'s method is the benchmark for such techniques, but it is not enough that such technology is available. For such techniques to be of use to researchers they must be widely adopted. One limitation, at least for small or impecunious institutions is that the technique relies on expensive third-party hardware and requires considerable expertise. Cheaper and more self-contained solutions could be more useful in the short term before the technology used by Mara et al. [2007] become affordable.

6.3.3.2 Other Methods

There are, however, many other systems available that are cheaper and perhaps more suited to the study and dissemination of vase-paintings. For example, QTVR (Quick Time Virtual Reality) object movies are obtained by stitching a variety of images of an artefact together into a composite model which may be viewed from multiple angles. In it's simplest form, QTVR models are akin to scrubbing a movie of an artefact rotating around a central axis - allowing the viewer to view the artefact from any angle around that axis. Such software is cheap - in some cases free - although getting good results often requires a lot of work. More expensive solutions involving mechanical devices for translating and rotating the camera automatically can generate very elaborate 3D models that may be viewed from any angle. Although the QTVR format admits hotspots to be defined on the object movies, without a lot of work, these may be defined on only a single frame.

Another technique that is widely used in Mayan pottery, but almost never used in the study of black-figure, is rollout photography. The rollout photograph, sometimes called peripheral photograph, is an unwrapped image of the surface of a cylindrical object achieved by photographing only a narrow vertical slice of the object while the object rotates about its vertical axis. When the object has rotated the full 360 degrees a complete rollout scan of its surface is obtained. In this sense, peripheral photography is very similar to a technique frequently used in the study of aurorae, called keograms.[11] This particular method of visualisation quickly allows scientists to spot anomalies in the space-time signature of the aurora. For example, Hough et al. [1992] discovered superfast auroral propagation using space-time diagrams whose signatures were not easily discernible in conventional imaging. It is quite possible that in an analogous way, rollout photographs may also be good visualisation techniques for art historians to recognise idiosyncrasies in a painter's style. More importantly, while the two attribution methods based on paintings presented in this vase were based on small features on the vase' surface, rollout photographs can be used as the basis of studies into the larger compositional features in a painting, such as the artists' use of space and the proportions of the whole figures.

The methods described above are both considerably better than the current methodology for the study of the surface paintings of vases. However,

[11] after the Inuit word *keoeit* (aurora).

neither of them are particularly useful for the study of the potting style. For this, there are a number of methods available that may be used for accurately scanning the geometry of the vases. On this basis, more detailed studies of vase shape may be carried out. Of these, laser profiling is possibly the simplest. This is an extremely accurate method for obtaining three-dimensional point-clouds of objects. The technique is used very widely and is equally applicable to large and small scale modeling: it has been used to model everything from heart valves to glaciers and indeed has been used extensively in archaeology, primarily in the image capture of rock-painting and also in the study of Mayan vases. There are a wide variety of laser-profiling techniques. In general a laser illuminates a point or set of points (such as a line) on an object that is then photographed by one or more cameras and the geometry calculated. Laser profiling may be implemented in a variety of different ways using different camera angles, and different ways of translating or rotating the laser beam or sheet so that most of the surface of the vase can be photographed in this manner.

6.3.3.3 The Way Forward

The techniques suggested here are far from exhaustive and are not mutually exclusive. It is possible that laser profiling, rollout photography, object movies and close-up images are all integrated into a composite data format that can be used for web publication of polychrome ceramics. For example, while studies relating to vase-shape and pottery recognition may be conducted on the basis of laser-scans, studies of the painting may be conducted on the object movies, rollout photographs and close-ups. A possible subject for future research is the development of methods of constructing three-dimensional modelling, perhaps integrating the written descriptions that usually accompany photographs (as in the *CVA*) as hyperlinks to the areas of interest. Of course, research is one thing and the practice of museums and other institutions that house these artefacts is another.

Despite the availability of third party systems, research into methods suited specifically for vase-paintings could be an important area of research, but one that is properly the remit of museums and academic institutions. Such systems can easily exploit the rotational symmetry of the vase about the vertical axis for both simplifying the process of capture and the compression of the point-clouds. Moreover, for the amount of data required to properly exploit techniques such as those presented in this thesis, there must be very widespread adoption of the 3D or pseudo 3D publication of ceramics. This applies not only to new excavations, but more importantly, the existing database (of over 100 000 vases) needs to be updated. This being achieved requires cheap and quick methods to be available since both expensive and time consuming techniques are unlikely to be implemented by the majority of museums. Therefore a possible area of future research is into cheap, easy, user-friendly methods finely tailored to the capture and publication of Attic vases.

6.3.4 Summary

Current technology allows for many exciting ways of presenting archaeological data to the interested user. Adapting these technologies to suit the needs of scholars of Greek pottery may be a fruitful area for future research. Issues that should be dealt with include the role of new web-based metadata tagging systems in the description of the data and issues of delivery such as whether to use a single repository or multiple sites. Most importantly, research into improving the quality of the visual data is badly needed. 3D imaging techniques should replace the standard of 2D Black-and-White photographs that are currently the norm. In particular, there is some scope, at the moment, for developing affordable 3-D imaging systems that are designed to meet the needs of scholars studying Greek vases. While state of the art systems certainly exist, more momentum is required before such systems become the norm for disseminating images of these artefacts.

6.4 Conclusion

The methods used in this dissertation represent only small subset of the vast number of possible methods of computer-aided attribution, and there are many more that may be developed in the future. This chapter, the purpose of which was to conclude the study by locating it within the broader discipline, provided brief guidelines for future research into this field. Two particular areas came under focus. First, it was suggested that future research into the field be conducted systematically in order to build a more complete picture of the attribution process with the ultimate aim of establishing a theory of attribution that holds true for all artistic media. To this end not only should findings from art history and the cognitive disciplines be incorporated into classifier design, but a culture should be developed whereby results are verified independently and negative results published so that more a complex picture can be painted of the way in which attribution should be carried out. The second part of the chapter concerned the process by which vases should be published, given that studies such as this suffer greatly from both scarcity and poor quality of the data. To some extent the situation may be ameliorated by the wide-spread adoption of publishing artefacts using good quality three dimensional images rather than the standard black and white photography.

Appendix A

The Princeton Painter and his *Oeuvre*

A.1 Introduction

By the middle of the 6th century, a new set of themes had emerged in Attic black-figure, and a new spirit. There were at least two paths open to the new vase-painter. One was the path of the masters, in which great effort was poured into the development of new themes, and the faithful rendering of anatomical details. This was the path taken by Exekias and his followers. This spirit would find its true home in the red-figure art that would replace black-figure as the medium of choice for the fine craftsman in the last quarter of the century. It is the second path that the Princeton Painter took. These artists valued production over art and often their work is cruder and more rushed as though aesthetic concerns were secondary to getting the job done. Such lesser masters are seldom studied and until the 1980's (Böhr [1982], Maxmin [1979] wrote monographs on the Swinger and the Painter of Berlin 1686 respectively) received little attention from scholars who favoured works by the finer painters like Exekias[Mommsen, 1997, Mackay, 1981, Technau, 1936] and the Amasis Painter [Karouzou, 1956, Robertson et al., 1987, von Bothmer, 1985]. Yet it is along the path of the lesser masters that the fate of black-figure lay as mass produced trinkets became the norm. This appendix introduces this personality and the problems associated with the study of his works and offers a very tentative chronology of the works attributed by Beazley or explicitly endorsed by him. There have been a number of attributions to the Princeton Painter by scholars after Beazley that are not in Beazley's lists. No attempt is made in this study to place these in this chronology as this thesis has been strictly based on Beazley's attributions.

A.2 Biography

There is very little external biographical information on Athenian vase-painters and what Classicists believe about them has been inferred from a handful of inscriptions and dedications. Very few of the painters signed their names and the vast majority of names that we use to denote them were invented by John Beazley. The Princeton Painter is no exception - he is named after one of his vases housed in the Princeton Art Museum (Princeton Y169). There are only two meaningful inscriptions on vases attributed to the Princeton Painter - only one on a vase attributed by Beazley. The Beazley example is on Bonn 365: *Onetorides Kalos* "Onetorides is beautiful". These kalos inscriptions are often interpreted as indications of affection of older men to youths they were courting. The Athenian tradition allowed older men to pursue romantic affairs with young aristocratic youths before they had reached maturity. The relationship had the older act as mentor and while there was clearly some sexual element in the relationship, it was primarily one of master and pupil. This inscription appears on other vases, all related to Exekias and his followers. This point will be picked up later in this appendix (A.5).

There is a further clue, although not on a vase attributed by Beazley, but which is interesting enough for it to deserve brief mention. A fragmentary krater attributed to the Princeton Painter by Ellen Page has an *epoiese* inscription.[1] The inscription is difficult to decipher and has received some attention. Moore offers IECHEKREIDES EPOIESEN while acknowledging suggestions of IEKSEKLEIDES EPOIESEN by Martin Robertson and IEXEKREIDES EPOIESEN by Dietrich von Bothmer [Moore, 1975]. Boegehold [1983] argues that it may well read WEKSEKLEIDES EPOIESE, which is not an Attic name. If Boegehold is correct then this would suggest a non-Attic origin for whoever potted the krater. While von Bothmer claims that the Princeton Painter did not paint any 2 vases by a single potter, there is evidence that this may not be the case (Chapter 6) and it is possible that the non-Attic name is the name of the potter of a number of Princeton Painter vases. Even if this is the case and the attribution is correct, however, it is unlikely that the name is that of the painter since the signature suggests a literate painter, but the nonsense inscriptions on one of his vases[2] suggests otherwise. It is more likely that the Bonn inscription and the potter inscriptions were incised by the potter or some other hand than that of the painter.

A.3 Style

The Princeton Painter's style is distinctive in a number of ways. His signature, at least from the Morellian perspective, has already been discussed

[1] Samothrace 71.1014A; 71.1072; 65.1060.

[2] Tarquinia 624 (M5).

in 3.2 but will be summarised here. The Princeton Painter's rendering of minor anatomical details appears to be guided by a sense of economy. The incisions he uses to render anatomical details are often single lines placed in such a way as to convey the essence of the anatomy rather than as a close study. Thus, for example, his ears are rendered with a single "C" shape, with a hook, or with an "S" shape which may be contrasted with Exekias who, as Mackay [1981, pp.58-62] shows, strives to render this anatomical detail more and more faithfully as his career progresses(figure A.1).[3] The Princeton Painter's mouths are perhaps the most economical of his forms - they are usually rendered with just a single short line. Maxmin [1979, p.174] has noticed these features of the Princeton Painter and compares them with later works of the Painter of Berlin 1686, whom she believes was influenced, towards the end of his career, by the former. Perhaps as an indication of the rush to finish his vases, the Princeton Painter sometimes omits to render weapons in the hands of his warriors, even when they appear poised to strike, such as in figureA.11.

The Princeton Painter is an enigma when it comes to his subject matter. While he is often very conservative in his compositions, he balances this out with some of the most unusual and difficult compositions to decipher. Thus, while a large number of his subjects are stock themes like Theseus and the Minotaur, battles, arming scenes, and the Iliupersis, there are other templates which are very unusual. For example, Herakles under the tripod (Munich 1378 figure A.7), the sale of oil (Brussels R279 figure A.6), and a seated man between winged females (Princeton 168, figure A.4) are extremely rare in black figure. Other templates are not unique but are rendered in a slightly unusual way, such as the duplicated scene of the recover of Helen (Once Peak collection, figure A.17). In addition, there are truly innovative compositions, such as the birth of Athena on the obverse and the moment after the birth, with Athena on Zeus' lap on the reverse of the same vase (Geneva MF 154, figure A.5). The moment after the birth of Athena is an uncommon motif and this is the only vase of which I am aware that has both scenes, appearing as though the painter was trying to render two separate frames of the same scene. Another innovative template is the appearance of the Delian triad (Apollo, Artemis and Leto) surrounded by gods - the common template is the two surrounded by human onlookers. While the scene is not unique, it appears as though the Princeton Painter's example is the earliest. Another unique scene is that of two women in a biga (a two wheeled chariot) drawn by two winged horses - a scene that is also unique

[3]The Princeton ear is a close-up of Zeus on the obverse of Geneva MF 154 (R2) and the Exekian example is from Vatican 344 *ABV* 145.13 *Para* 60 *Add*^{2} 40.

(a) (b)

Figure A.1: An ear by Exekias and one by the Princeton Painter. (a) is an image of Exekias's ears which reveal similar structure to a real human ear. The Princeton Painter's offering (b) is a shorthand that is anatomically incorrect by conveys the right effect.

among extant vases, and one that is not easily interpreted. The list could continue - roughly half of the Princeton Painter's vase include a composition that is unusual in some respect. Thus, the Princeton Painter's compositions are either so unusual as to border on the bizarre, or they are stock standard, perhaps even copied from other works such as those of Group E (a theme picked up in A.5.3.1).

The subsidiary decoration of the Princeton Painter is unadventurous and is quite typical of many his contemporaries. The shoulder friezes on belly amphorae are generally lotus and palmette festoons (examples from different painters are shown in figure A.3 B, C & D and figure A.2 shows a labeled drawing of a single element - the palmette is the top part of the element), or arcaded lotus buds (a Swing painter example in figure A.3 A), sometimes with black dots in the intersections. The palmettes are short, as is typical of the period, but they are also more round and articulated than those of his contemporaries (compare figures A.3 A, B & C), and quite unlike the elegant splayed separate leaves of Exekias (figureA.3 E). In some cases, the lotuses are buds rather than flowers (figure A.3 F), an uncommon decoration that

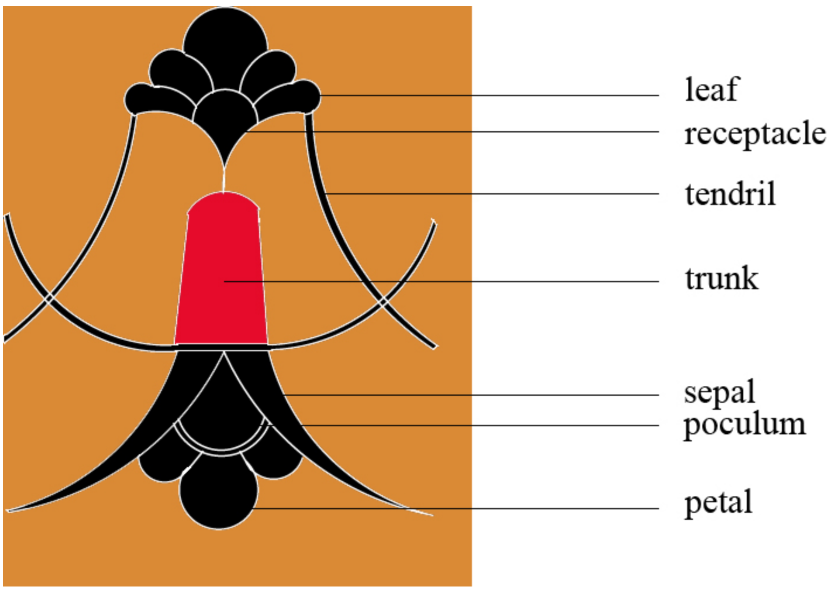


Figure A.2: A single element of the lotus and palmette labeled. The labelling follows, for the most part, Mackay [1981, p.372], save that the term "leaf" has been used where Mackay uses petal on the palmette.

the Princeton Painter shares with the Swinger.[4] The sepals of the bud are painted different colours. The trunk of the lotus is often painted red and this is usually untidy with the red paint seldom neatly covering the whole area of the trunk (for example the reverse of Basel BS 427 (figure A.10(b)). On neck-amphorae, it is the standard opposed lotus and palmette chains. Quite often the palmette leaves are alternating colour - usually red and black. Neck-amphorae also have the typical sub-handle friezes of spirals with palmettes at the intersections. These are sometimes very tightly packed (for example on St. Petersburg 162 (figure A.16) which is unusual for the period, since sub-handle palmettes are usually splayed by the 530s, such as is evident on Cincinnati 1884.213 (figure A.9). In addition, two of the Princeton Painter's neck-amphorae have animal processions on the predella (the decorated frieze below the scene). Both these motifs are quite old fashioned and illustrate the Princeton Painter's archaising tendency.

[4]Rhodes 1346 B **M1**, Bonn 45 A **EM4** and New York 56.171.9 B **M10**.

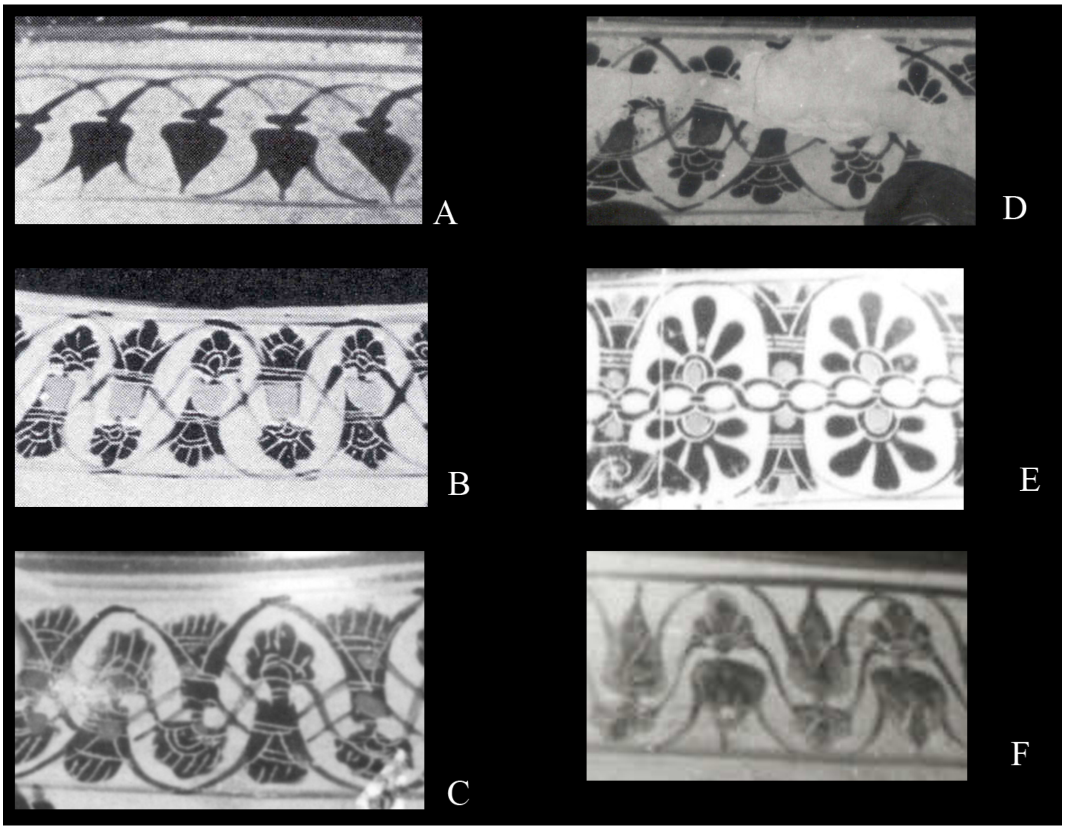


Figure A.3: Subsidiary neck decorations from Belly amphorae. A. Inverted arcaded lotus buds by the Swing Painter (Vatican G37 B) B. Lotus and palmette festoon by the Swing Painter (London B195 A C. Lotus and palmette festoon by the Painter of Berlin 1686 (Louvre F3) D. Lotus and palmette festoon by the Princeton Painter (Basel BS 427 A). Opposed lotus and palmette chain by Exekias (Faina 186) F. Lotus bud and palmette festoon by the Princeton Painter (Basel BS 427 B).

Finally, in terms of shape, the Princeton Painter favours type B amphora, and among his amphorae attributed by Beazley, there exist no type A amphorae. In addition to these shapes, the Princeton Painter also has an oinochoe, 3 hydriae, 5 neck amphorae and 3 pseudo-panathanaics. The reason for the abundance of type B amphorae may reflect the preference of the painter or, given that there are so few vases attributed to him, may simply be due to chance. It is also quite possible that the Princeton Painter's career did not span that long. Indeed, there is little evidence of much artistic development in his corpus.

A.4 Related Artists

A.4.1 The Princeton Group

The Princeton Group is a large group of disparate painters that Beazley considered close enough to the work of the Princeton Painter to name them after him. These include the following: The Manner of the Princeton Painter, Group of St. Petersburg 1469, Group of Munich 1379 and the Group of Munich 1393. Of these, the closest to our man is obviously the Manner of the Princeton Painter which, I imagine, contains works that are close enough to be the work of the Princeton Painter, but with which Beazley had sufficient doubt to place them in a different Group. Since the Princeton Group is not the work of a single painter, there is no point in explaining the points of similarity - this heterogeneous collection of painters are similar enough to the Princeton Painter that most of the Princeton Painter's stylistic features are represented in this corpus, but on none of these paintings is there sufficient evidence of his personality for a firm attribution.

It should be pointed out that some of the works from the Princeton Group may well be the work of other known painters. Jody Maxmin [1979, pp.191-197] has assigned all the paintings from the Painter of Munich 1379 to the Painter of Berlin 1686. In addition, Boardman has also attributed a vases[5] from Beazley's manner of the Princeton Painter to the Painter of Berlin 1686[6] and notes[7] that the shape and painting of a vase near the painter of Berlin 1686[8] and the painting on the underside of the foot is identical to that on a vase in the manner of the Princeton Painter.[9] An obvious question is whether any vases in the Princeton Group are actually the output of the Princeton Painter. The answer may lie in a subsequent research project. The converse may also be asked - how many of the works attributed to the Princeton Painter are the work of one of the other painters in the Princeton Group. My own feeling is that most of Beazley's attributions are safe, but if the Princeton Group represent imitators of our man, it does appear strange that after Beazley, there have been far more attributions by other scholars to the Princeton Painter than to his followers and one may well ask how many of these attributions rightly belong elsewhere in the Princeton Group than in the corpus of the painter himself.

[5] Oxford 1965.141 *ABV* 299.1

[6] *CVA* Great Britain 14 Oxford 3 (Ashmolean) p.19

[7] *CVA* Great Britain 14 Oxford 3 (Ashmolean) p.20

[8] *ABV* 301 middle.

[9] *ABV* 300.6.

A.4.2 The Swing Painter

The artist most closely associated with the Princeton Painter by most people's reckoning was the Swing Painter. Beazley considered the Swing Painter to be a student of the Princeton Painter (*ABV* 297). Elke Böhr, who has written the only published monograph on the Swing Painter, did not agree with this assessment. She does notice that there is considerable similarity with the Princeton Painter's work, she sees this as evidence of collaboration rather than a master-student relationship [Böhr, 1982, p.53]. Some of the points of contact between the painters are discussed below.

First, the Princeton Painter and the Swing Painter both employ a very economical style. Just as our man, the Swing painter uses approximations, if not short-hand, for anatomical details. These include the single line for the mouth and the odd incision to indicate anatomy. This Swing Painter also shares the Princeton Painter's habit of forgetting to render the weapons in some fight scenes [Smith, 1945, p.469]. There are many similarities in the drapery, with the himation wrapped around the male figure being rendered with a row of diagonal bands across the figure's body. This feature is a slight modification of a more archaic expression of this mode of dress and in many ways reveals a lack of awareness of the more naturalistic expressions evident in the works of many of their contemporaries.

There are also many themes that are shared by these painters, including some that appear in Group E and some that are just plain unusual. Like the Princeton Painter, the Swing painter occasionally renders scenes that defy interpretation, and many of these are scenes that bear some compositional and stylistic similarity with the Princeton Painter's uninterpretable scenes. For example, the reverse of Princeton Y168 (figure A.4) by the Princeton Painter is almost the same template as that on the Swing Painter's Vatican G37 (figure A.8). A. final point of contact between these two painters is the manner in which the Princeton Painter and the Swing Painter render ancillary details such as animals, subsidiary decorations and accessories. In particular, an analysis of horse-details from Moore [1971] reveals that of the specific forms associated with equine anatomy, the Swing Painter and Princeton Painter share more in common with each other than with any other painters (although for some reason, they both appear to share many features with Group E and Exekias).

A.4.3 The Painter of Berlin 1686

The painter of Berlin 1686 was linked with the Princeton Painter by Beazley, and there are indeed numerous similarities between these two painters. Much like the Swing Painter, the Painter of Berlin 1686 also uses economical shorthand on some of his vases. Maxmin [1979, Ch8] believes these to be the product of a later phase of the Painter of Berlin 1686's career when he appears to be influenced by whom she believes are more fashionable contemporaries, the Princeton Painter and the Swing Painter. On the other hand, I suspect economic motives may have been more of a concern than fashion since other contemporaries like Exekias and the like don't seem particularly keen on the slapdash style of these painters.

The painter of Berlin 1686, unlike the Princeton Painter and Group E, has the ability to render not only well-thought out paintings, but also fine detail - something that is absent from all of the Princeton Painters work. Some similarities that are shared between the two is that they use similar lotus and palmette festoons and the arcaded lotus buds, though the detail is somewhat different. In particular, while the Princeton Painter has somewhat neater and more articulated palmette leaves, the overall impression is of a slightly more untidy subsidiary decoration on the Princeton Painter's works.

A.4.4 Group E

The Princeton Painter's works have some points of contact with the work of Group E, particularly in terms of composition and subject matter. Particularly the vases that I have labeled 'early' show considerable influence from Group E in terms of composition. The template of Herakles struggling under the weight of the tripod on Munich 1378 (figure A.7) mirrors a composition in Group E that is otherwise unique in extant black-figure (discussed in more detail in A.5.3.1). Furthermore, some of the less adventurous of the Princeton Painter's compositions are similar to those of Group E. The similarities are not limited to painting style, however, there are similarities in the pottery as well. In particular, the special neck-amphora with a broad shoulder allowing a separate scene to be decorated there is popular only in Group E, The Painter of London 174 (related to Group E), Exekias and the Princeton Painter, all but one (Boulogne 4) of whose (non-panathanaic) neck-amphora are of this kind.

In addition to Group E proper, it is worth mentioning one point of contact with the loftier work of Exekias and his followers. Bonn 356 has the kalos inscription dedicated to Onetorides. This occurs elsewhere only in the works of Exekias (5 vases) and his followers. In terms of shape, a rather rare form of neck-amphora of the type described above, with the broad shoulder, is used by the Princeton Painter in one Beazley attribution (St. Petersburg 162, figure A.16). This form has three figured friezes - one on the shoulder, the main frieze, and an animal predella. This particular type seems to be used almost exclusively by Exekias and the Painter of London 174[Moore, 2007, p.26]. In addition to St. Petersburg 162, an amphora in New York attributed to the Princeton Painter by von Bothmer[10] is of this kind.

Whether these points of contact with Group E and Exekias are because there was a working relationship between the painter is not certain since it may also reflect just how wide the influence of Group E and Exekias was. While the point may not have been noticed by other scholars, Webster [1972, p.27] in his analysis of the groups and workshops in 6th century Athens, lists the Princeton Painter tentatively as belonging to the same tradition as Exekias, Group E and the Lysippides painter (Exekias' pupil by most accounts).

A.5 Chronology

One of the reasons for studying individual painters apart from the nuanced understanding of the iconography that such detailed studies offer, is that they increase our understanding of the chronological development of the tradition, allowing us finer dating of the pieces. This is not a trivial issue since, if we use the paintings as cultural evidence for the era, there could be a big difference between vases that are painted 20 years apart. For example, this is the difference between the reign of Peisistratos (546-527 BCE) and the birth of democracy in Athens (509 BCE). Thus, a key issue in the study of individual painters has traditionally been the establishment of a chronology of the painter's work based on a study of the painter's internal development. Although there have been a number of articles devoted to the Princeton Painter or individual works by the artist, there has as yet been no attempt to deal with the artist's internal chronology. This point will be dealt with after a brief exposition of how chronologies are traditionally established.

A.5.1 Techniques for Establishing internal chronologies

The establishment of a chronology is usually based on a number of important pieces of evidence. First, the rendering of the human form and its drapery changes over the course of the sixth century. Particularly, over the second half of the century, painters become comfortable with $\frac{3}{4}$ views. While the painters at the middle of the century frequently make perspective errors and have difficulty faithfully rendering the musculature, this is eventually

[10]This is not a Beazley attribution, but the typical features of the Princeton Painter's style described in Chapter 1 are present and in the author's opinion the attribution is safe.

perfected, at least in as much as the strict tradition will allow. This development, to a large extent, parallels similar development in relief sculpture. The drapery become more and more naturalistic at the same time, and there is a greater willingness on the part of the painters to render the folds of the drapery in an attempt to give the illusion of volume and depth. At first this is simply achieved by rendering simple bands and panels to indicate folds in the drapery, but these bands and panels become freer and narrower and eventually three-dimensional folds create the illusion of body. Towards the end of the century these folds become stylised, rendered haphazardly and without much thought as mass-production starts to over-ride aesthetic concerns. A second area that can reveal chronological clues is the rendering of subsidiary decorations. Possibly the most noticeable trend in this regard is for the palmette fronds to become elongated and separately splayed. A final tendency over the course of the sixth century is the tendency of vases to become more and more slender over the course of the century, a practice that has already been discussed in Chapter five. This means that if a painter painted exclusively the wares of a single potter then it is likely that their chronological development will track a tendency of the vases to become more slender.

The most important evidence, however, has to be inferred from the output of the individual artist. Through art historical intuition, the scholar has to gain an impression of what motivates the artist and use this to plot a trajectory for the artist's creative career. This development needn't be in a straight line, for example, as is illustrated by Maxmin [1979]'s chronology of the Painter of the Berlin 1686 which starts with a period of technical excellence (Ch1, 2), a transitional phase of abbreviating forms (Ch4), a phase of exuberance (Ch4) and two final phases in which his style become closer and closer to those of the Princeton Group and the Swinger (Ch5, 7). In some cases, such as that of Exekias, the painter's motivations may be more simple: more and more realistic production. Of course, realism is not the only artistic motivation for a painter. There may well be complex and nuanced issues that motivate the artist including design, fine detail, showmanship, composition, and the dictates of the market. Whatever the case, getting some feeling for the aims of the artist does help the scholar understand the trajectory his art takes over the course of his career, and consequently allows the scholar to locate new examples within this development.

A.5.2 Difficulties with the Princeton Painter's Corpus

However, the task of establishing a chronology for the illusive personality of the Princeton Painter has proved intimidating to scholars far greater than the author. Dietrich von Bothmer, in attributing an impressive Neck-amphora in Geneva to the Princeton Painter, studiously avoids the task of contextualising the vase within the development of the Painter's career [Chamay and von Bothmer, 1987]. Instead, he points to various features that would on most paintings be chronologically significant, but which the Princeton Painter uses with too much freedom to be of much use. He comments, on the task of placing the Geneva vase within the chronological sequence of the Princeton Painter's works

> "...this is not quite so easy as it would appear, for some painters do not develop in a straight line with clearly recognisable chronological landmarks, and others show a deplorable tendency to be inordinately inconsistent in the quality of their works."[Chamay and von Bothmer, 1987]

Von Bothmer then proceeds to illustrate the Princeton Painter's inconsistency in interpretation of contemporary fashions that make a mockery of the usual chronological markers: The Princeton Painter is not interested in anatomical details, and on almost all his vases there appear the familiar approximations of detail rather any evidence of an attempt at realism; he is innovative with his subsidiary decoration, and thus resists standardising it; he does not appear to have worked for a single potter, so one cannot use vase shape as a reliable indicator of development; and finally, as Boardman [1974, p.63] states , "the Princeton Painter and his ilk show awareness rather than understanding of new developments, as in the representation of drapery".

A.5.3 An approach to the chronology of the Princeton Painter

The Princeton Painter is clearly active from around the middle of the sixth century to somewhere before the middle of the second half of the century. Traditional wisdom holds that the Painter of Berlin 1686 is earlier than the Princeton Painter who is earlier than the Swing Painter. Böhr [1982, p.56] gives 540-520 as the probable dates for the Swing Painter, so we should expect slightly earlier starting date for the Princeton Painter. Judging from the Princeton Painter's lack of awareness of developments in red-figure, a date earlier than 530 is probably likely for his later works. Thus, a time spanning roughly 545-530 is plausible and will be the starting point for this study.

To begin the investigation proper, we should seek external points in the chronological development whereby certain individual vases may be linked with certain external sources such as other vases or historical people. The most significant piece of evidence is an amphora in Bonn (figure A.13) depicting a man on a horse riding next to a void horse lead by a naked youth. The vase is quite unusual in that there is no subsidiary decoration, and the composition is very stark and minimalist for the Princeton Painter, who likes to fill out areas of empty space instead of using his space to greater effect. This superb composition and balance perhaps led Beazley to comment that the work was remarkably good for the Princeton Painter.(*ABV*) This is all the more significant if one considers Beazley's comment at the beginning of the chapter on "Other Pot Painters" into which category our man falls: "Those who are reading the book through may be inclined to skip them..."[*ABV* 296] The sparseness of the design recalls a particular type of austere amphora called horseman amphora that may have been tributes to the deceased. However, the sense of balance surely reflects a mature artist, or at least, an artist in his prime.

Another clue on the Bonn amphora is the inscription *Onetorides Kalos* "Onetorides is handsome". Any date later than 530 would beg the question of how long a youth could be *kalos* considering the list of archons records the name Onetorides in 526/7 BC. A man close to thirty, the minimum age for archon, should surely not have been the object of an older man's affection. Assigning an earliest date is difficult. The inscription appears elsewhere and seems to be connected with Exekias because all the vases that have the inscription are by Exekias or his successors.[11] Clement [1955, pp.9-10] has summarised the difficulties with using Onetorides for dating. To some extent it depends on what Exekian chronology one uses. Nevertheless, using Mackay [1981]'s chronology, it is likely that the earliest of these is later than 540. A different view is held by Webster [1972, pp.65-66] who gives a date range for *Onetorides Kalos* of between 560 and 530. He does not explain which of these vases could date back further than 550 but, given that it cannot be the vases related to the Lysippides Painter, it must either be Berlin 1720 of Exekias, or our own vase in Bonn. Of course, Webster has an agenda for extending the age of the vase so far back and that is to prove the thesis that kalos inscriptions are not messages of love but patronage - a point addressed by Robinson [1975] and otherwise not widely accepted by Classical scholars. A date before 540 for Berlin 1720 is extremely unlikely, as Mackay [1981]

[11]Exekias: London B210 *ABV* 144.7 *Para* 60 *Add*^{2} 39; Berlin 1720 *ABV* 143.1 *Para* 59 *Add*^{2} 39; Vatican 344 *ABV* 145.13 *Para* 60 *Add*^{2} 40; Athens, Agora AP1044 *ABV* 145.19 *Para* 60 *Add*^{2} 40; Near Exekias: Barcelona 4500A (fragment) *ABV* 148 *Add*^{2} 41; Manner of the Lysippides Painter:(Mastos Painter) New York 14.105.10 *ABV* 261.37; Related to the Lysippides Painter: St. Petersburg *ABV* 264.2; 3-Line Group: Villa Giulia *ABV* *693.8bis.*

points out. Perhaps Webster sees the Bonn vase as one of the horseman amphorae popular before 550, but the Princeton Painter's archaisms should be considered when coming to such a conclusion. The forms on the vase, however, are typical of those that are much later in the Princeton Painter's career - a point picked up later.

There is little else to suggest the vase is earlier than 540 and without further evidence to the contrary, it should be considered roughly contemporary with other "*Onetorides Kalos*" vases. The bounds are still loose, but, Webster's comments aside, they place our vase somewhere between 540 and 530. Nevertheless, this vase must be considered much later than the amphora that will be discussed next and which provides our external evidence for the early phase of the Princeton Painter's work

A.5.3.1 Early Vases

A second vase provides another external chronological marker. Munich 1378 (figure fig:Munich 1378) depicting Herakles struggling under the weight of the tripod, the interpretation of which is uncertain, seems to be very early in the Princeton Painter's career. Beazley interprets the obverse of the vase as "Herakles hoisting his prize tripod". While Herakles is often depicted with a tripod, this is almost always within the context of his fight with Apollo over the Delphic tripod [12] - and this is clearly not a scene of Herakles and the Delphic tripod. In fact, most scenes of Herakles and the Delphic tripod only appear on Attic vases after the erection of the Siphnian treasury at Delphi in 525 BC, the eastern pediment of which depicts this scene [von Bothmer, 1977]. Tripods were used as prizes in Athletic events, and it is likely that Beazley was correct. The template of this vase is an almost identical copy of the template of two Group E amphora (figure A.14), one in Rome (both sides) and one in Copenhagen (reverse) that show a naked man carrying a tripod to right, while flanked by two naked men on either side of him. A number of compositional similarities between both sides of the these vases and the obverse of the Munich amphora suggest that the Princeton painter's vase was influence by the two group E pieces (or, less likely, vice verse). In the first instance, on all three vases the victor is flanked by two supporters on either side. On the Group E amphorae, these supporters are naked men, while Herakles, on the Munich amphora, is flanked by three naked youths.

[12]The story is that Herakles, in a fit of madness induced by Hera, kills a guest of his and has to purify himself. He seeks advice from the oracle at Delphi, but the priestess is so disgusted with his actions (he had, some time earlier, sought the oracle's advice to free himself from the pollution of killing his wife and children) that she refuses to grant him an oracle. Herakles steals the tripod and a fight with Apollo ensues.

Furthermore, on the Copenhagen vase, the two supporters on the right hold fillets in their right hands (which are at their sides), as do the youths on the right of Herakles on the Princeton painter's amphora. Even more significant is the way the respective artists have rendered the victor carrying the tripod. On both scenes on the two Group E vases and on the relevant scene on our vase, the victor moves to the right, his head bowed down and body hunched and almost kneeling under the weight of the device. In order to balance, his right hand holds the leg of the tripod behind him about half-way down the leg's length, and his left hand grips the leg in front of him near the top of the device. The stance of these central figures is quite different from the stances of the victors on the other vases, who tend to stand upright. The pose of the figures on these three vases, bowing down under the weight of their prize, bear little resemblance to any other depictions of people carrying tripods. Certainly, in the depictions of Herakles carrying the tripod, for example, he does so with ease, on many occasions managing to fight off Apollo with his spare hand. Given the uniqueness of the template and the similarities, the Group E vases and the Princeton Painter vases are related, and may be roughly contemporary. The Group E vases are from roughly 550-545, and this period may well be right for the Munich Amphora.

Having established two chronological points in the painter's career, it is possible to examine the elements that characterise Munich 1378 and on this basis find vases that may be contemporary. First, the forms that characterise the Princeton Painter's work (as described in 3.2) do not appear with any regularity and there does not appear to be any consistency in the way in which details such as musculature are rendered from figure to figure. The effect of the forms on the Munich amphora is unappealing and they are not good proxies for the real anatomy. More importantly, the forms are inconsistent - many of the Princeton Painter's other paintings have more consistency in the manner in which these anatomical details are rendered. It is tempting to see these forms as representing some experimentation on the part of the painter who is trying to establish a technique that is both visually appealing and easy to render. In this regard it is worth pointing out that the forms the Princeton Painter uses on his other vases are in no way more difficult to render than the unattractive examples on the Munich amphora and moreover there is not much other evidence that the Munich amphora was executed in a rush. It is quite unlikely then that a painter who has developed an appealing and quick way to render shorthand such as the C-shaped knee and the S shaped ear would switch to the haphazard collection of forms just out of expediency. Furthermore, we should intuitively expect a painter to become more consistent, rather than less consistent, in the rendering of details that he has practised for years and for which he does not have to reflect. Another piece of evidence pointing at an early date is the difficulty with which the painter renders $\frac{3}{4}$ views. The painter renders the figure on the immediate right of Herakles, for example, with his nipples visible to the viewer, suggesting that the arm that attaches to the right (our right) shoulder is the figure's left arm, yet the hand on this arm is occluded by the figure's body as though it were the right hand. Indeed, the right arm has been occluded by the figure's body and only the right hand is visible on the left (our left) of the figure. Another manifestation of the painter's struggle with this view is that the nipples of all his figures are rendered at different heights.

I tentatively group together the other vases in the corpus that share similar style and on which forms are rendered haphazardly. These probably represent the earliest phase in the Painters career. They are Princeton Y168, Munich 1378 and New York 53.11.1. There are a number of other features that these vases have in common. First, these vases have very unusual templates which defy interpretation - they, for the most part, are unique in black figure. In addition, these vases are also very static, both in terms of composition and in terms of the manner in which the figures and the drapery are rendered. Each of them has strong vertical emphasis, and the templates are not particularly adventurous, being mostly 2 bystanders on each side of a central group, with the notable exception of the obverse of New York 53.11.1. By contrast, non-descript bystanders are surprisingly rare in the Princeton Painter's later work, occurring again on the reverses of 4 vases I have classed as early-middle, and then only on two other vases in his corpus. Even in scenes in which non-descript bystanders would be expected, the Princeton Painter avoids them or uses gods to fill these positions. Two examples from his later works illustrate this: on both sides of his birth of Athena vase in Geneva (figure A.5), he frames the scene with Apollo kitharoidos on the left, and on both sides, he frames the scene on the right with Poseidon. And on London B212 (figure A.18), the reverse depicts the Delian Triad (Apollo between Artemis and Leto), a scene in which bystanders usually frame the three Delian deities, but which here have Hermes and Poseidon framing the scene instead. The early vases also have in common a perspective error resulting from the painter's marginally successful attempts at the $\frac{3}{4}$ view: the figure immediately to the right of the seated figure on the obverse of New York 53.11.1 and the figure on the extreme right of both sides of Princeton Y 168.

A.5.3.2 Early Middle Period

In addition to these early vessels there are four other vases that may be somewhat earlier than the rest of the Princeton Painter's corpus, and which I refer to as the early-middle vases. The rendering of the forms is still haphazard, but the established Princeton Painter forms are also present. These vases, in addition, appear to be pre-occupied with the Trojan war. The vases are Basel BS 427 (figure A.10), Madrid 10925(EM 2), Once Peek (figure A.17) and Bonn 45(figure A.12) Like the early vases, there are scenes that defy interpretation because the painter appears to be unaware of the tradition. The drapery on these vases is less static than the early vases. The fabrics are still rendered either in broad panels or bands and contain little texture (except for dot-rosettes and fish-scales) or added white. However, a small degree of freedom is claimed by the drapery, which now hugs the bodies of some of the figures. Furthermore, the hems of the stationary figures are not always flat, but sometimes wavy and sometimes diagonal. There is still no attempt at rendering any folds, however. In terms of composition, the painter has become considerably more adventurous. The battle scenes are vivid and full of movement, and even the stationary scenes have some degree of movement in the central group. However, these vases also show some difficulty with the $\frac{3}{4}$ view, although the results are somewhat better than the earlier vases. For example, the warriors in Basel B427 (figure A.10)have narrow chest compared with other vases by the same painter. A feature that occurs quite often in the Princeton Painter's corpus is the single spiral on the breast-armour. However, on the Basel amphora it is rendered differently - on other Princeton Painter vases, there are either two spirals or one spiral and a large area of blank space so that the chest is the normal size (such as in figure A.11).

An interesting issue with these vases is the very unusual scene that appears on two vases from this group, Madrid 10925 and Basel BS 427 . This is a scene of men in battle with a statue of Athena to left with spear raised. Beneath Athena's shield is a warrior who has collapsed (Basel) or kneels (Madrid) to right but looks back at a warrior approaching from left to right and engaging in the fallen warrior. The compositional template is actually taken from another popular scene in black-figure, the rape of Kassandra (an example of which is illustrated in figure A.15). The story is that during the sacking of Troy, the hero Ajax raped the priestess Kassandra while she clung to the statue of Athena in the Palladion for protection. The typical rape of Kassandra scene has Kassandra kneeling beneath the shield of Athena's statue (that faces left) and Ajax approaching from left to right attacks Kassandra. It is clear that both these vases are inspired by this template. Vian has suggested that the scene represents the gigantomachy[13] which is one of the few scenes in black-figure in which goddesses like Athena appear in battle. There are two problems with this interpretation. First, it rests on the figure of Athena actually being part of the battle, though her stance, her gaze, and the fact that the figure directly approaching her is clearly not engaging with her, make it difficult to see her as engaging in battle. Secondly, Athena almost always fights to right in gigantomachies (victors almost always approach from the left in black-figure) and there is the question of why she alone of the gods should be depicted - in gigantomachies there are usually multiple gods. She is surely a statute and the similarity with the rape of Kassandra scene suggest the scene is from the Trojan war. Apart from this, it is difficult to suggest a specific moment from this event.

The other vases from this group are an amphora once in the Peek collection (figure A.17) but which has since been sold to a private buyer, and an amphora in Bonn (figure A.12). The Bonn amphora shows the death of Priam on the obverse and a fight on the reverse, while the Peek vase has a scene of Helen being recovered, along with a second woman, possibly a war bride, on the obverse, and a departure scene on the reverse.[14]

A.5.3.3 Mature Vases

From this point it becomes very difficult to make educated guesses about the Princeton Painter's chronology because most of the paintings do have all the Princeton Painters established forms and because some of these vases have a combination of features that are associated with art from the 530's but which are often mixed with archaic elements that may well have been consciously rendered. The established forms, such as the "S " shaped ear and the "C" shaped knee (which are explained in 3.2 are expressed on Bonn 365 - the vase we dated to between 540 and 530 earlier in the chapter (A.5.3). The painter's art matures in a number of ways. First, the forms that are characteristic of his style generally appear on most of the vases. Secondly, his rendering of stock scenes are quite standard, particularly the fight scenes.

However, when the painter does deviate from the standard representation it does not appear that this is due to lack of understanding of the traditional conventions. Instead these vases appear to be contemplative and very deliberate. While it is impossible to be certain, it is possible that when the painter had enough money to relax the pace, at times, he could afford to forsake profit to produce one or two indulgent pieces. The catalogue at the end of this appendix lists these contemplative pieces separately as the final

[13]The battle between the gods and the giants.

[14]That is a scene of a warrior setting out into battle, often in a chariot.

phase of the painter's career, but this by no means implies they are all later. It is possible (and from stylistic considerations likely in the author's opinion) that many are contemporary with some of his more rushed works, but congregate toward the end of the career.

An example of such a work is London B212 (figure A.18) which shows the Delian triad between gods rather than bystanders is employed by some later painters as well. This suggests that either the Princeton Painter's composition influenced others, or that the Princeton Painter was painting a scene that was current at the time, but no other early examples of which have survived. Further illustration that the painter understood the tradition is that on the shoulder of the London vase appears a full account of the fight between Herakles and Kyknos, in a scene reminiscent of the Hesiodic poem, *The Shield of Herakles*. The poem is a detailed account of the battle of Herakles and Kyknos which features a full-fledged battle complete with thundering chariots that appear on our vase as well.[15] Furthermore the centrepiece of the shield of Herakles as it is described by Hesiod is a scene of Apollo entertaining the gods with his lyre.[16] It is thus tempting to see parallels with the scene on the reverse of this vase with that on his shoulder.

Another contemplative effort is the study of the marketplace in Brussels R279 (figure A.6) depicting the sale of oil or more likely wine.[17] The scene is quite unusual in black-figure since studies of everyday life are more common in red-figure. Particularly interesting are the subtle elements of realism in a very stylised genre: there are big differences in apparel: some wear what appear to be sacks (slaves) and their beards are rough and shaggy resembling those of satyrs; and the portly lady (possibly indicating a middle-aged matron rather than the typical maiden) is very unusual, as is the fact that she has incisions on her cheeks, probably to indicate creases or wrinkles.

The vases from this mature period make up the vast majority of the vases attributed by Beazley to this painter. There are some generalisations we can make about the style of the painter during this period, although there are counter-examples to all of these. First, the drapery gains increasing volume. The bands indicating folds of the himatia gradually become thinner and more wavy, and there are even examples of three-dimensional folds such as in the Orvieto neck-amphora (figure A.19) showing that the painter was aware of some of the changes that were going on around him. Of these

[15] For a discussion of the relationship between the poem and the scenes on vase paintings, see Shapiro [1984].

[16] [Evelyn-White, 1964, ln.201-203]

[17] Beazley suggests the sale of oil, although the reverse scene shows an oinochoe on the wall and Immerwahr [1992] points out that one of the sellers dispenses his wares from a wineskin.

vases, the Bonn vase is perhaps the best because it not only has some fine draftsmanship. Archaism abound, however, and the subsidiary decorations remain conservative. For example, he still uses animal processions on the predellae of his neck-amphorae and the sub-handle ornaments on the same are typically palmettes of the old variety (for example on St. Petersburg 162: figure A.16) - only the neck-amphora in Cincinnati (figure A.9) has the new splayed variety that characterised most of the work of his contemporaries, and the neck-amphora in Orvieto has something in-between.

A.6 Conclusion

It is difficult to come to any firm conclusions about the Princeton Painter's chronology because the area is a minefield. A tentative chronology has been provided here, but this must be considered a very tentative chronology, since the Princeton Painter's archaising makes it difficult to determine what is really old and what is meant to simply convey the impression of archaic grandeur. In addition, the external points that are used to peg the two vases around which the chronology is developed are fuzzy. The bounds on the age of the Bonn amphora are loose and the relationship between the two Group E amphora and Munich 1378 does not have to imply that they are contemporary.

Moreover, the purpose of this appendix is to present the painter whose work forms the basis of this study and to offer a more detailed discussion for the reader whose interest is peaked by the more circumspect descriptions in the body of the thesis. The Princeton Painter's style is elusive and enigmatic and deserves the full attention of a dedicated monograph that also examines the relationship between him and the works of the Princeton Group. Chapter 5 provides some evidence that the Princeton Painter may have painted for a few potters. If this is the case then it is tempting to believe that we may indeed have his signature on the fragmentary krater in Samothrace which, if Boegehold is correct, is not Attic. Although it is unlikely that the signature is his, a foreign metic is perfectly consistent with the chronology established here - a young foreigner trying to get to grips with the alien tradition which he often misinterprets, and who gradually becomes master himself. His eye is never on technical mastery, but probably on profit which his quick shorthand must have allowed him to pursue. One may hope that the more contemplative works later in his career like London B212 and Bonn 365 represent the output of a man who had achieved his aim of relative wealth and was able to spend more time than was profitable to an art that he had always relegated to the dictates of the market. Of course this is simply romantic speculation. Given the paucity of information available to us we should be cautious about taking a speculative chronology such as this as fact. Computer-aided attribution is in its infancy and it will probably take a synergy of human and machine to get to the bottom of the questions that are tentatively explored in this appendix.

A.7 Catalogue of Princeton Painter Vases

The catalogue of Princeton Painter vases is listed according to the categories established in the section on chronology (A.5). The author does not have good-quality images of all the vases, and some of the analyses in this studies were conducted on the basis of images from the Beazley archive which are watermarked. Instead of providing a set of plates, the catalogue also lists the Beazley archive number for each vase, which the interested reader may consult. For those vases for which the Beazley Archive does not have images and the author does, the catalogue provides a link to the image. Within each date range, vases are not listed according to chronology, with the caveat that within these phases, the chronology is even less certain. Moreover, I have listed Munich 1378 as the first and Bonn 365 as the last because they are the "anchors", but this does not mean that they are the earliest or latest in their phase respectively.

A.7.1 Early

E1 Munich 1378 *ABV* 299.17; *Add*^{2} 78; BN 320416
E2 Princeton 168 *ABV* 299.19; *Add*^{2} 78; BN 320418
E3 New York 53.11.1 *ABV* 298.5; *Add*^{2} 78; BN 320404

A.7.2 Early Middle

EM1 Basel 427 ;*Para* 130.14bis; *Add*^{2} 78
EM2 Madrid 10925 *ABV* 298.11; *Add*^{2} 78; BN 320410
EM3 Once Peek *ABV* 298.12; *Para* 129; BN 320411
EM4 Bonn 45 *ABV* 299.16; BN 320415

A.7.3 Mature

M1 Rhodes 1346 *ABV* 298.7; BN 320406
M2 Cahn 313 *Para* 130.6ter
M3 Once Paris *ABV* 298. 12
M4 Villa Giulia 910 *ABV* 298.9; BN 320408
M5 Tarquinia 624 *Para* 130.15bis; *Add*^{2} 78;
M6 Villa Guilia 20872-3; *ABV*
M7 Cambridge GR 1.1889 *ABV* 298.10; *Add*^{2} 78; BN 320409
M8 Naples Stg 144 *ABV* 298.13; BN 320412
M9 Roman Market *ABV* 298. 14; BN 320413
M10 New York 56.171.9 *ABV* 299.15; *Para* 129; *Add*^{2} 7; BN 320414

M11 Marburg 1009 *ABV* 299.23; BN320422
M12 New York 23.160.92 *ABV* 299.24; *Para* 130; BN 320423
M13 Cincinnati 1884.213 *ABV* 692.4bis; BN 306595
M14 Orvieto *ABV* 298. 4; BN 320403

A.7.4 Contemplative

R1 Swiss Private *Para* 130.5bis ;Add^2 78
R2 Geneva MF 154 *ABV* 298.18; *Para* 130; Add^2 78; BN 320417
R3 Louvre F 217 *ABV* 298. 2; BN 320401
R4 Boulogne 4 *ABV* 298. 3; BN 320402
R5 St. Petersburg 162 *Para* 130.1bis ;Add^2 78
R6 Basel Ka 411 *ABV* 299.25; 451; *Para* 130; Add^2 78; BN 320424
R7 Princeton 169 *ABV* 299. 19; BN 320405
R8 Brussels R279 *ABV* 299.20; Add^2 78; BN 320419
R9 London B212 *ABV* 297.1; *Para* 129; Add^2 78; BN 320400
R10 Bonn 365 *ABV* 299.21, 672; BN 320420

A.8 Images of Vase Scenes

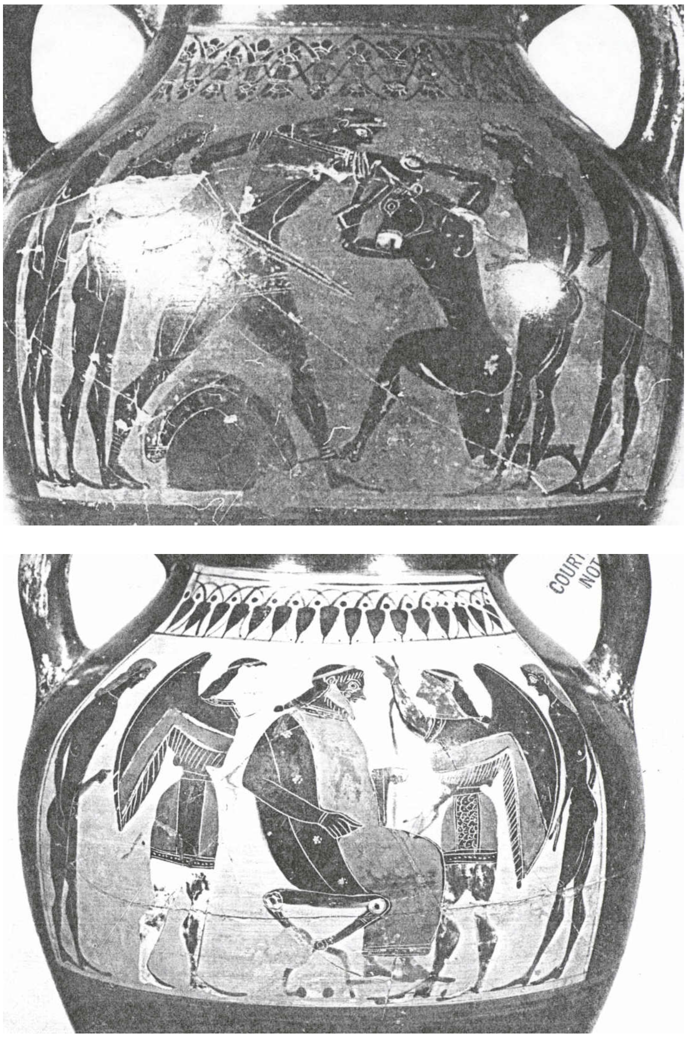


Figure A.4: Princeton 168 (E2) (a) Theseus and the Minotaur and (b) possibly the pre-birth of Athena .

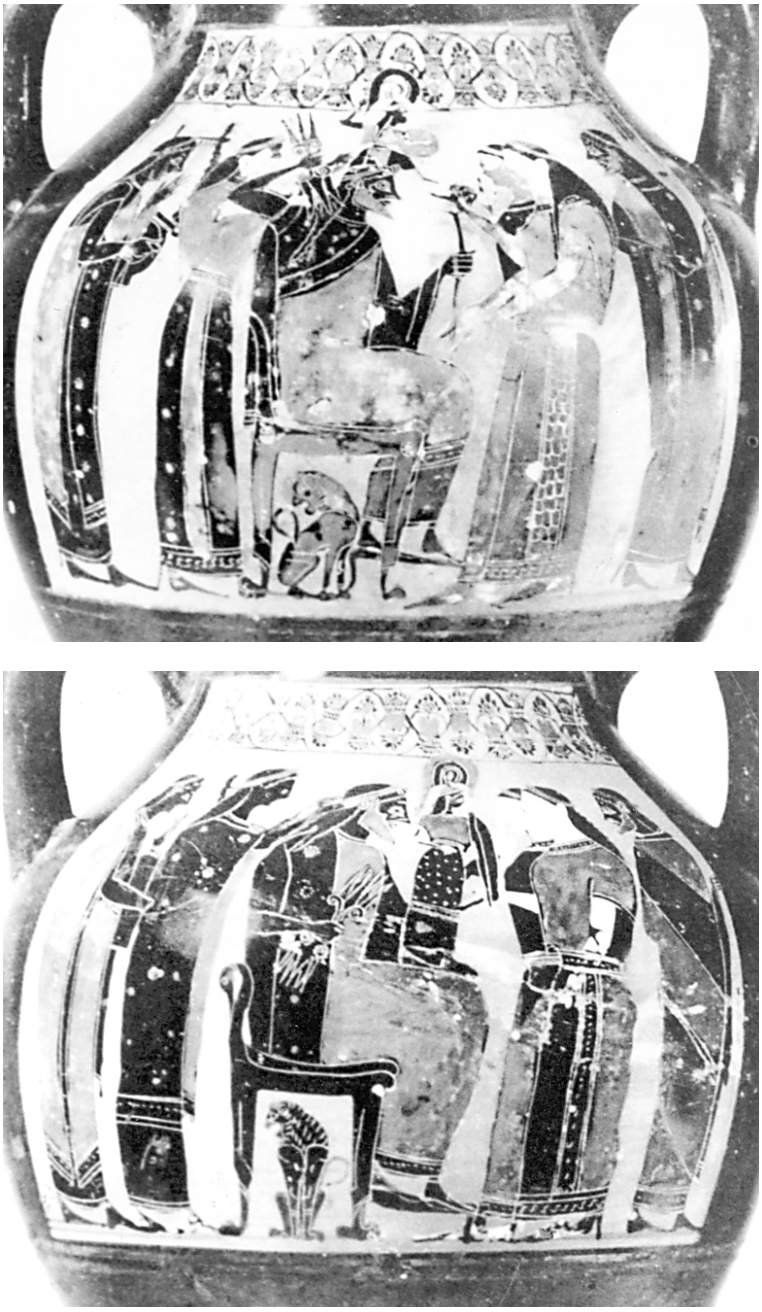


Figure A.5: Geneva MF 154 (R2) The birth (a) and post-birth (b) of Athena. Photograph courtesy of Anne Mackay

Figure A.6: Brussels R279 obverse **R8** The sale of oil/wine. Photocopy from the Beazley archive.

Figure A.7: Munich 1378 obverse (E1) Herakles carrying away the prize Tripod. Photograph from *CVA* Germany 3 Munich 1 (Museum Antike Kleinkunst) pl. 11.4.

Figure A.8: A seated man between two winged females. Vativan G37 by the Swing painter. Photograph from Böhr [1982, pl. 19].

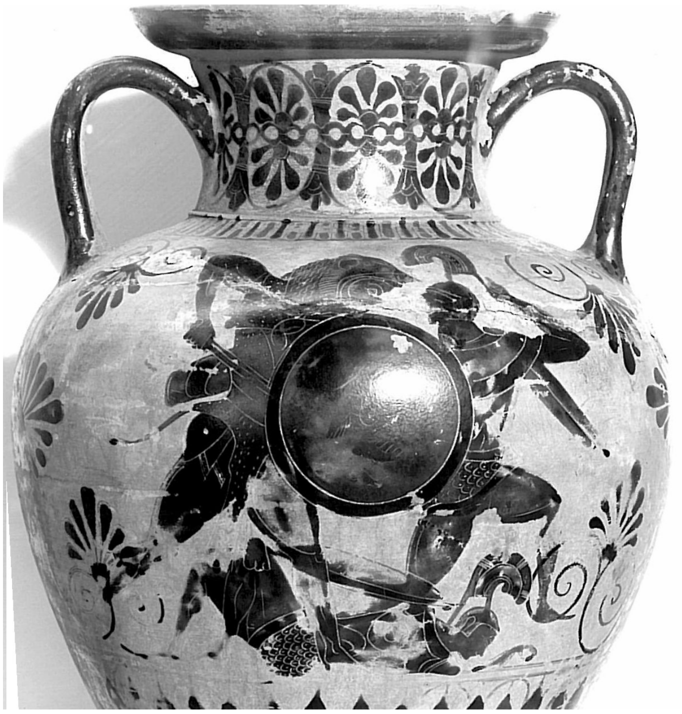


Figure A.9: Obverse of Cincinnati 1884.213 **M13** Herakles fighting the Amazons. Photograph courtesy of the Cincinnati Art Museum.

(a)

(b)

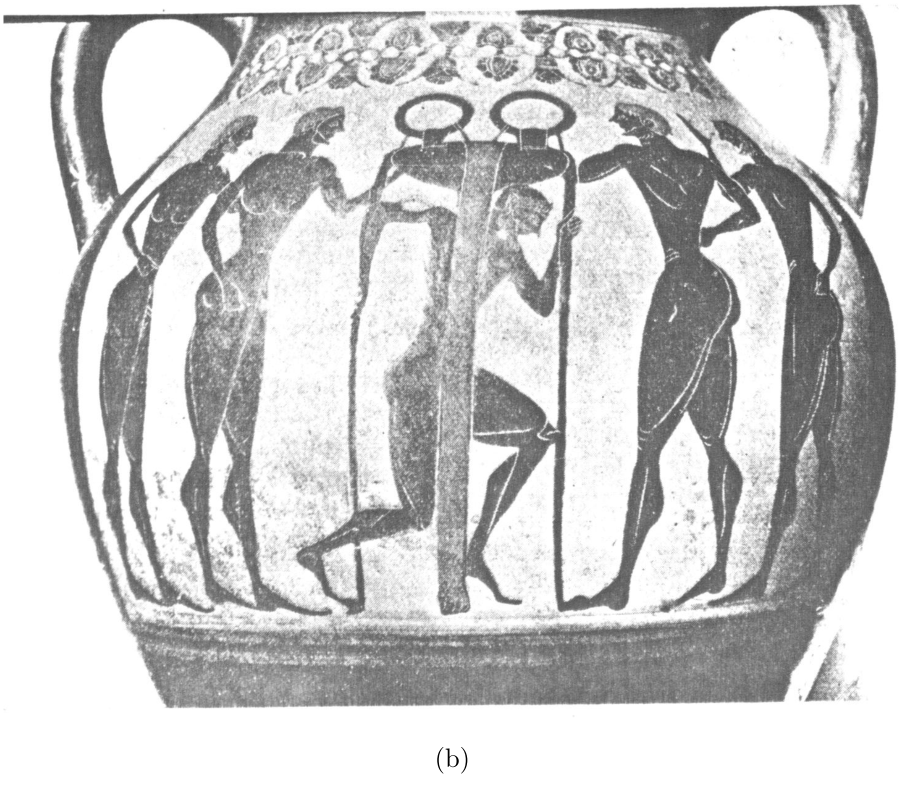
Figure A.10: Basel 427 **EM1**: (a) Gigantomachy or the Rape of Kassandra and (b) Theseus and the Minotaur. Photograph courtesy of the Antikenmuseum Basel und Sammlung Ludwig.

Figure A.11: Detail of the obverse of New York 56.171.9 **M10** Notice that there is a single spiral on the breastplate, but just in front, there is empty space allowing the full sized $\frac{3}{4}$ view chest. Photograph courtesy of Anne Mackay.

Figure A.12: Bonn 45 obverse **EM4** The sack of Troy. Photograph courtesy of the Akademisches Kunstmuseum, Bonn.

Figure A.13: Bonn365 obverse **R10** A youth on a horse lead by a squire. Photograph courtesy of the Akademisches Kunstmuseum, Bonn.

(a)



(b)

Figure A.14: Obverse of (a) Copenhagen 109 (reverse is the same) and (b) Rome, Guglielmi: Both show an athlete struggling under the prize tripod. Photographs courtesy of Anne Mackay.

Figure A.15: Berlin 1698. Obverse: The Rape of Kassandra (top) and revers: Theseus and the Minotaur (bottom). Photograph courtesy of Anne Mackay.

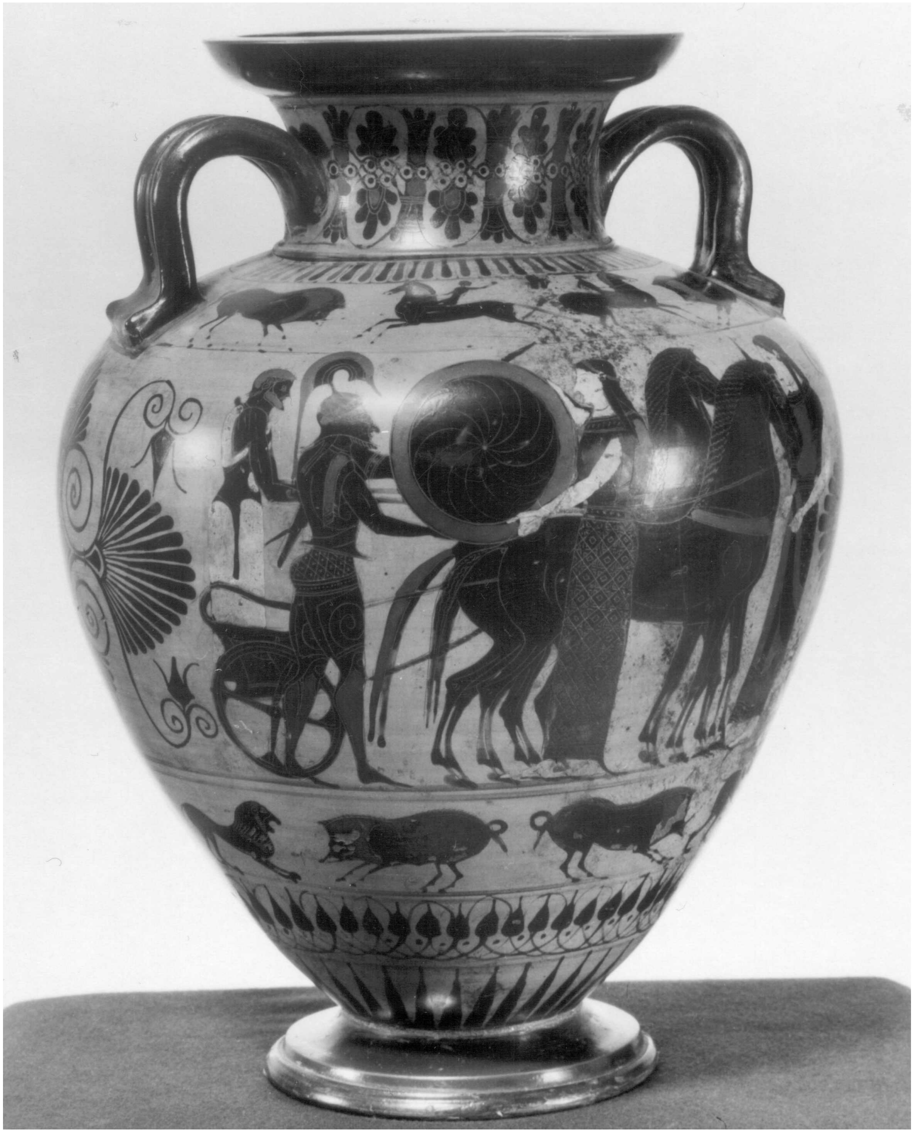


Figure A.16: St. Petersburg 162 **R5**: The animal predella is visible beneath the scene. The sub-handle ornament reveals very tightly bunched palmette leaves which are somewhat anachronistic. Photograph courtesy of the Hermitage Museum.

Figure A.17: Once Peek collection obverse **EM3** Recovery of Helen. Photocopy from the Beazley Archive.

Figure A.18: London B212 obverse **R9** Apollo plays the lyre between among the gods. Photograph courtesy of Anne Mackay.

Figure A.19: Orvieto obverse **M14** Fight. The vase displays some awareness of current developments including the splayed folds of the chitoniskos (short tunic) on the warrior, and the slightly splayed palmettes beneath the handle. Photocopy from the Beazley archive.

Bibliography

H. Ackerman and J.-R. Gisler, editors. *Lexicon Iconographicum Mythologiae Classicae*. Artemis Verlag, Zurich and Munich, 1981-. 2, 12

K. Adler, M. Kampel, R. Kastler, M. Penz, R. Sablatnig, K. Schindler, and S. Tosovic. Computer aided classification of ceramics - achievements and problems. In *In Proceedings of 6th International Workshop on Archaeology and Computers*, pages 3–12, 2001. 36

N. Ahmed and K. R. Rao. *Orthogonal transforms for digital signal processing*. Springer Verlag, 1975. 181

M. Aksela and J. Laaksonen. Using diversity of errors for selecting members of a committee classifier. *Pattern Recognition*, 39:608–623, 2006. 85

C. Aras. Hindsite: A robust system for archaeological fragment re-assembly. Master's thesis, Division of Engineering, Brown University, 2007. 36

R. Arnheim. A review of proportion. *The Journal of Aesthetics and Art Criticism*, 14(1):44–57, 1955. 129

I. S. Ayelet Gilboa, Avshalom Karasik and U. Smilansky. Towards computerised typology and classification of ceramics. *Journal of Archaeological Science*, (31):681–694, 2004. 35

D. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981. 115

S. Barrat and S. Tabbone. A progressive learning method for symbols recognition. In *SAC '07: Proceedings of the 2007 ACM symposium on Applied computing*, pages 627–631, 2007. 180

J. Beazley. *Attic Black-figure Vase-painters*. Oxford University Press, Oxford, 1957. 2

J. Beazley. *Attic Red-figure Vase-Painters*. Clarendon Press, Oxford, second edition, 1968. 2, 12

J. Beazley. Amasea. *Journal of Hellenic Studies*, 51(2):256–285, 1931. 171

J. Beazley. Groups of mid-sixth-century black-figure. *Transactions of the British School at Athens*, 32:17–18, 1932. 30

J. Beazley. *Paralipomena: additions to Attic Black-Figure Vase-painters and to Attic Red-figure Vase-painters*. Oxford University Press, Oxford, 1971. 2, 12

B. K. Behl. Tradition in bronze. *Frontline*, 27(7), 2008. 197

B. Berenson. *Seeing and Knowing*. MacMillan, New York, 1954. 15, 18

T. Bin, L. Siwei, and H. Hua. High performance face recognition system by creating virtual sample. In *IEEE international Conference on Neural Networks and Signal Processing, Nanjing, China*, pages 972–975, 2003. 100

P. Biro, T. Ebersole, M. Felder, P. M. I Jensen, N. Smith, and T. von Liptak. An initial investigation into aspects of preservation potential of the Bradshaw rock-art system, Kimberley, Northwestern Australia. *Antiquity*, 75(288):257–258, 2001. 33

P.-P. Biro. Biro fine art restoration and forensics, web page (last accessed january 21 2009), 2008. URL `http://www.birofineartrestoration.com`. 33

G. Bishop. *Classification of Greek Pottery Shapes and Schools using Image Retrieval Techniques*. PhD thesis, Pace University, 2006. 36

H. Bloesch. *Formen Attischer Schalen von Exekias bis zum Ende ses Strengen Stils*. Benteli, Berne, 1940. 41, 161, 173, 176

H. Bloesch. Stout and slender in the late Archaic period. *Journal of Hellenic Studies*, 71:29–39, 1951. 162

J. Boardman. *Athenian Black Figure Vases*. Thames and Hudson, London, 1974. 141, 234

J. Boardman. The Amasis Painter. *Journal of Hellenic Studies*, 78:1–3, 1958. 171

J. Boardman. Herakles, Peisistratos and sons. *Revue Archaeologique*, pages 57–72, 1972. 3

J. Boardman. Herakles, Peisistratos and Eleusis. *Journal of Hellenic Studies*, 95:1–12, 1975. 3

J. Boardman. Herakles, Peisistratos and the unconvinced. *Journal of Hellenic Studies*, 109:158–159, 1989. 3

J. Boardman. The new Attic potter. *Oxford Journal of Archaeology*, 2: 247–248, 1983. 30, 163

J. Boardman. Crafty arts: review of Vickers and Gill (1994). *Classical Review*, 46(1):123–126, 1996. 19, 22

J. Boardman. *The History of Vase-painting*. Thames and Hudson, London, 2006. 15, 17

A. Boegehold. A new Attic black-figure potter. *American Journal of Archaeology*, 87:89–90, 1983. 30, 163, 224

E. Böhr. *Der Shaukelmaeler*. Philipp von Zabern, 1982. 30, 38, 96, 141, 223, 230, 234, 249

U. Braga-Neto. Fads and fallacies in the name of small-sample microarray classification. *IEEE Signal Processing Magazine*, 24(1):91–99, 2007. 58, 61

U. Brago-Neto and E. Dougherty. Bolstered error estimation. *Pattern Recognition*, 37:1267–1281, 2004. 81

D. Brain and G. Webb. On the effect of datset size and on bias and variance in classification learning. In *Proceedings of the Fourth Australian Knowledge Acquisition Workshop*, pages 117–178, 1999. 61

L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996. 85

L. Breiman. Arcing classifiers. *Annals of Statistics*, 26(3):801–824, 1998. 87

F. Brommer. *Vasenlisten zur griechischen Heldensage*. NG Elwert Verlag, Munich, third edition, 1973. 2, 12

S. T. Buckland, C. Burnham, K. P. Burnham, and N. H. Augustin. Model selection: an integral part of inference. *Biometrics*, 53:603–618, 1997. 76

T. Carpenter, editor. *Beazley Addenda: additional references to **ABV**, **ARV**^{2} and **Paralipomena***. Oxford University Press, Oxford, second edition, 1989. 2, 12

J. Chamay and D. von Bothmer. Ajex et Cassandre par le peintre de Princeton. *Antike Kunst*, 30:58–68, 1987. 30, 31, 162, 234

K. J. Cherkauer. Human expert level performance on a scientific image analysis task by a system using combined artificial neural networks. In *Working Notes, Integrating Multiple Learned Models for Improving and Scaling Machine Learning Algorithms*, pages 15–21, 1996. 86

J. Churchill, J. Stanis, C. Press, M. Kushelev, and W. Greenough. Is procedural memory relatively spared from age effects? *Neurobiology of Aging*, 24(6):883–892, 2003. 92

D. Clark. Matrix analysis and archaeology with particular reference to British beaker pottery. *Proceedings of the Prehistoric Society*, 28:371–72, 1962. 35

P. Clement. Geryon and others in Los Angeles. *Hesperia*, 24(1):1–24, 1955. 235

P. Conner and H. Jackson. *A Catalogue of the Greek Vases in the Collection of the University of Melbourne*. MacMillian Art Publications, 2000. 3, 216

R. M. Cook. Amasis mepoiesen. *Journal of Hellenic Studies*, 68:148, 1948. 171

R. M. Cook. Pots and Pisistran progaganda. *Journal of Hellenic Studies*, 107:167–169, 1987a. 3

R. M. Cook. Epoiesen on Greek Vases. *Journal of Hellenic Studies*, 91: 137–138, 1971. 20

R. M. Cook. Artful crafts: a commentary. *Journal of Hellenic Studies*, 107: 169–171, 1987b. 21, 22

E. Cox. *The Poems of Sappho*. Williams & Norgate, New York, 1925. 22

J. N. Davie. Theseus the king in fifth century Athens. *Greece & Rome*, 29 (1):25–35, 1982. 3

T. G. Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10:1895–1923, 1998. 106, 107

T. G. Dietterich. Ensemble learning. In M. A. Arbib, editor, *Handbook of Brain Theory and Neural Networks, Second Edition*, pages 405–408, 2002. 85

J. E. Doran and F. R. Hodson. A digital computer analysis of palaeolithic flint assemblages. *Nature (London)*, 210:688–689, 1966. 35

R. Duda and P. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972. 115

R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley Interscience, New York, 2000. 71

P. Durham. *Image Processing and Hypermedia Tools for Archaeological Classification*. PhD thesis, University of Southampton, 1996. 36, 98, 213

P. Durham, P. Lewis, and S. Shennan. Artefact matching and retrieval using the generalised Hough transform. In *Computer Applications and Quantitative Methods in Archaeology 1993*, pages 25–30, 1995. 35

P. Durham, P. Lewis, and S. Shennon. Image processing strategies for artefact classification. *Proceedings of Computer Applications in Archaeology 1995: Anaelectica Praehistorica*, 28:235–240, 1996. 35

R. Elia. review of Vickers and Gill (1993). *American Journal of Archaeology*, 100(2):422–423, 1996. 25

H. G. Evelyn-White. *Hesiod, the Homeric Hymns and Homerica*. Heinemann & Harvard University Press, London and Cambridge MA, 1964. 241

T. Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27: 861–874, 2006. 83

D. Fortenberry. Single greaves in the later Helladic period. *American Journal of Archaeology*, 95(4):623–627, 1991. 98

M. Friedländer. *From van Eyck to Breugal (tr. M Kay)*. Phaidon, London, third edition, 1969. 15

M. Friedländer. *On Art and Connoisseurship (tr. T Borenius)*. Beacon Hill, Boston, 1960. 9, 129

J. Friedman. On bias, variance and the 0/1-loss, and the curse of dimensionality. *Data Mining and Knowledge Discovery*, 1:55–77, 1997. 59

D. Gill and C. Chippendale. Material and intellectual consequences of esteem for Cycladic figurines. *American Journal of Archaeology*, 97(4):601–659, 1993. 29

J. Gips. *Shape Grammars and Their Uses: artificial perception, shape generation and computer aesthetics*. PhD thesis, University of California Los Angeles, 1975. 32

J. Gips and G. Stiny. An evaluation of Palladian plans. *Environment and Planning B: Planning and Design*, 5:199–206, 1978. 32

J. Gips and G. Stiny. Shape grammars and the generative specification of painting and sculpture. In *InfoProc71*, volume 8, pages 213–220, 1972. 32

R. C. Gonzalez and R. E. Woods. *Digital Image Processing.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001. 117

C. Goutte. Note on free lunches and cross-validation. *Neural Computation*, 9:1245–1249, 1997. 67

M. Guy and E. Tonkin. Folksonomies: tidying up the tags? *D-Lib Magazine*, 12(1), January 2006. 217

N. Hall and S. Laflin. A computer aided design technique for pottery profiles. In *Computer Applications in Archaeology 1984*, pages 178–188, 1984. 35

J. Hambidge. *Dynamic Symmetry: The Greek Vase.* Yale University Press, New Haven, 1920. 199

H. M. Hansen. Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology*, 58:321–334, 1959. 197

T. Hastie and G. James. Generalizations of the bias/variance decomposition for prediction error, 1997. 59

G. Hersey and R. Freedman. *Possible Palladian Villas.* MIT Press, Cambridge MA, 1992. 32

S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979. 166

G. Hough, M. J. Kosch, and M. J. W. Scourfield. First observations of superfast auroral waves. *Geophysical Research Letters*, 19(24):2433–2435, 1992. 219

H. Immerwahr. New wine in ancient wineskins: the evidence from Attic vases. *Hesperia*, 61(1):121–132, 1992. 241

G. James. *Majority Vote Classifiers: theory and applications.* PhD thesis, Stanford University, 1998. 59

D. Jenkinson. The elicitation of probabilities - a review of the statistical literature. University of Sheffield: Bayesian Elicitation of Experts' Probabilities (BEEP) working paper, 2005. 207

S. Jiang and T. Huang. Categorizing traditional Chinese painting images. In *Advances in Multimedia Information Processing: 5th Pacific Rim Conference on Multimedia*, pages 1–8, 2004. 34, 212

W. H. S. Jones. *The Natural History, volume VI.* Heinemann & Harvard University Press, London and Cambridge MA, 1951. 21, 22

K. Jones-Smith and H. Mathur. Fractal analysis: revisiting Pollock's drip paintings. *Nature*, 444:E10, 2006. 34

M. Kampel and R. Sablatnig. Rule-based system for archaeological pottery classification. *Pattern Recognition Letters*, (28):740–747, 2007. 36

M. Kampel, R. Sablatnig, and E. Costa. Classification of archaeological fragments using profile primitives. In *In Computer Vision, Computer Graphics and Photogrammetry a Common Viewpoint, Proceedings of the 25th Workshop of the Austrian Association for Pattern Recognition*, pages 151–158, 2001. 36

S. Karouzou. *The Amasis Painter.* Clarendon Press, Oxford, 1956. 171, 223

A. King. Whodunit? art sleuth uses fingerprints to put an artist's name to a painting. *Montreal Gazette*, October 12 2000. 33, 205

J. Kirsch and R. Kirsch. The anatomy of painting styles: description with computer rules. *Leonardo*, 21(4):437–444, 1988. 32

R. Kirsch. Photogrammetric reconstruction of petroglyphs. *American Indian Rock Art*, 23:177–182, 1997. 32, 33

R. Kirsch. Using computers to describe style. *American Indian Rock Art*, 22:153–160, 1998. 32

R. Kirsch and J. Kirsch. The structure of paintings: formal grammar and design. *Environment and Planning B: Planning and Design*, 13:163–176, 1986. 32

R. Kirsch, L. Cahn, L. C. Ray, and G. H. Urban. Experiments with processing pictorial information with a digital computer. In *Proceedings of the Eastern Joint Computer Conference,December 9-13, 1957*, pages 151–158, 1957. 32

T. Knight. Transformations of the meander motif on Greek Geometric pottery. *Design Computing*, 1:29–67, 1986. 32

T. Knight. Transformations of de Stijl art: the paintings of Georges Vantongerloo and Fritz Glarner. *Environment and Planning B: Planning and Design*, 16:51–98, 1989. 32

R. Kohavi and D. Wolpert. Bias plus variance decomposition for zero-one loss functions. In *Proceedings of the 12th International Conference on Machine Learning*, pages 313–321, 1996. 59, 61

W. Kong. Archaeological fragment reassembly using curve matching. Master's thesis, Division of Engineering, Brown University, 2002. 36

E. G. Kraeling. The evolution of the story of Jonah. In A. Caquot, editor, *Hommages á Andre Dupont-Sommer*, pages 305–18, 1971. 2

W. Kropatsch, M. Eder, and P. Kammerer. Finding strokes of the brush in portrait miniatures. In *Proceedings of the 19th ÖAGM and 1st SDRV Workshop, Maribor, 1995*, volume 81, pages 474–476, 1995. 34

A. Kuhl, L. K. C. Wöhler, and U. Kressel. Training of classifiers using virtual samples only. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 04)*, pages 418–421, 2004. 100

P. Lewis and K. Goodson. Images, databases and edge detection for archaeological object drawings. In *Computer Applications and Quantitative Methods in Archaeology*, pages 149–153, 1991. 35

P. Lewis, P. Durham, and S. Shennon. Artefact matching and retrieval using the generalised Hough transform. *Dept of Electronics and Computer Science: 1993 Research Journal*, pages 57–9, 1993. 35

F. Leymarie, D. Cooper, M. Joukowsky, B. Kimia, D. Laidlaw, D. Mumford, and E. Vote. The shape lab: new technology and software for archaeologists. In *Computing Archaeology for Understanding the Past (CAA 2000)*, pages 78–89, 2001. 36

J. Li and J. Wang. Studying digital imagery of ancient painting using mixture of stochastic models. *IEEE Transactions on Image Processing*, 13(3):340–353, 2004. 34, 212

T.-S. Lim, W.-Y. Loh, and Y.-S. Shih. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning*, 40(3):203–228, September 2000. 75

G. Liming, L. Hongjie, and J. Wilcock. The analysis of ancient Chinese pottery and porcelain shapes: a study of classical profiles from the Yangshao culture to the Qing dynasty using computerized profile data reduction, cluster analysis and fuzzy boundary discrimination. In *Computer Applications and Quantitative Methods in Archaeology 1989*, pages 363–374, 1989. 35

E. A. Mackay. *Exekias: A chronology of his extant works.* PhD thesis, University of Victoria, Wellington, 1981. 41, 141, 162, 173, 174, 175, 223, 225, 227, 235

E. A. Mackay. *Methodology in Vase-Profile Analysis*, volume 2. The J.Paul Getty Museum, Malibu, 1985. 173, 174

C. J. Mackie. The earliest Jason: what's in a name. *Greece and Rome*, 48 (1):1–17, 2001. 3

P. Main. The storage, retrieval and classification of artefact shapes. *Computer Applications in Archaeology*, pages 39–48, 1978. 35

P. Main. *A Method for the Computer Storage and Comparison of Outline Shapes of Archaeological Artefacts.* PhD thesis, National Council for Academic Awards, 1983. 35

C. Maiza and V. Gaildrat. Automatic classification of archaeological potsherds. In *8th International Conference on Computer Graphics and Artificial Intelligence, Limoges*, pages 135–148, 2005. 36

H. Mara, M. Kampel, and R. Sablatnig. Preprocessing of 3D-data for classification of archaelogical fragments in an automated system. In *26th Workshop of the Austrian Association for Pattern Recognition (OeAGM/AAPR*, pages 257–264, 2002. 36

H. Mara, E. Trinkl, P. Kammerer, and E. Zolda. 3D-acquisition and multispectral readings for documentation of polychrome ceramics in the antiquities collection of the kunsthistorisches museum vienna. In J. Trant and D. Bearman, editors, *International Cultural Heritage Informatics Meeting - ICHIM07: Proceedings*, 2007. 218, 219

C. Martindale. Peak shift, prototypicality and aesthetic preference. *Journal of Consciousness Studies*, 6(6-7):52–54, 1999. 197

J. Maxmin. A new amphora by the Painter of Berlin 1686. In E. Böhr and W. Martini, editors, *Stüdien zur Mythologie und Vasenmalerei : Festschrift für Konrad Schaünburg*, pages 35–40, 1986. 141

J. Maxmin. *The Painter of Berlin 1686.* PhD thesis, Oxford University, 1979. 141, 223, 225, 229, 231, 233

J. McBride. Archaeological fragment reassembly using curve matching. Master's thesis, Division of Engineering, Brown University, 2003. 36

M. McClellan. review of Vickers and Gill(1994). *Journal of Field Archaeology*, 23(6):390–393, 1996. 25

T. Melzer, P. Kammerer, and E. Zolda. Stroke detection of brush strokes in portrait miniatures using a semi-parametric and a model based approach. In *Proceedings of the 14th International Conference on Pattern Recognition*, pages 474–476, 1998. 34, 206, 212

P. Michaelsen, T. Ebersole, N. Smith, and P. Biro. Australian Ice Age rock art may depict Earth's oldest recordings of shamanistic rituals. *Mankind Quarterly*, 41(2):131–147, 2000. 33

H. Mommsen. *Exekias I.* Forschungen zur antiken Keramik, Meinz, 1997. 141, 223

M. B. Moore. The Princeton Painter in New York. *Metropolitan Museum Journal*, 42:21–56, 2007. 30, 31, 232

M. B. Moore. *Horses on Attic black-figured Greek Vases of the Archaic Period Ca. 620-480 BC.* PhD thesis, New York University, 1971. 199, 230

M. B. Moore. Attic black figure from Samothrace. *Hesperia*, 1932:238–250, 1975. 30, 162, 224

G. Morelli. *Italian Masters in German Collections (tr. LM Richter)*. Bell & Sons, London, 1883. 9

G. Morelli. *Italian Painters: Critical Studies of Their Work (tr. CJ Ffoulkes)*. J. Murray, London, 1892. 9

G. Morelli. *Die Werke italienischer Meister in den Galerien von Munchen, Dresden und Berlin: Ein kritischer Versuch.* E. A. Seemann, Leipzig, 1880. 9

G. Morelli. *Kunstkritische Studien uber italienische Malerei.* FA Brockhaus, Leipzig, 1890. 9, 15

E. Panofsky. The history of art as a humanistic discipline. In P. Alperson, editor, *The Philosophy of the Visual Arts*, pages 469–80, 1992. 8

T. Perneger. What's wrong with Bonferroni adjustments. *British Medical Journal*, 316:1236–1238, 1998. 155

R. Polikar. Ensemble based systems in decision making. *IEEE Circuits and Systems magazine*, 3rd quarter:21–45, 2006. 84

F. Provost, T. Fawcett, and R. Kohavi. The case against accuracy estimation for comparing induction algorithms. In *Proceedings of the Fifteenth International Conference on Machine Learning*, pages 445–453, 1998. 81

V. S. Ramachandran and W. Hirstein. The science of art: a neurological theory of aesthetic experience. *Journal of Consciousness Studies*, 6(6-7): 15–51, 1999. 52, 197, 198

G. Richter. *A Handbook of Greek art*. Phaidon Press, London, 1959. 7

M. Robertson, A. Stewart, J. Boardman, and W. Burkert. *Papers on the Amasis Painter and His World: colloquium sponsored by the Getty Center for the History of Art and the Humanities and symposium sponsored by the J. Paul Getty Museum*. J. Paul Getty Museum, Malibu, 1987. 171, 223

M. Robinson. Review of TBL webster, Potter and patron in Classical Athens. *Journal of Hellenic Studies*, 95:295–6, 1975. 235

S. Rotroff. Hellenistic Athenian pottery: Megarian bowls. *Current Anthropology*, 19(2):387–388, 1978. 22

R. Sablatnig, P. Kammerer, and E. Zolda. Heirarchical classification of paintings using face- and brush stroke models. In *Proceedings of the 14th International Conference on Pattern Recognition*, pages 172–174, 1998. 34, 206

R. Schapire. The boosting approach to machine learning: An overview. In *Nonlinear Estimation and Classification*, pages 1–23, 2003. 86

R. Schapire, Y. Freund, P. Bartlett, and W. S. Lee. Boosting the margins: A new explanation for the effectiveness of voting methods. *Annals of Statistics*, 26(5):1651–1686, 1988. 87

H. Shapiro. Herakles and Kyknos. *American Journal of Archaeology*, 88(4): 523–529, 1984. 241

S. Siegel. *Non-parametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York, 1956. 108

E. Simon. review of Vickers and Gill (1994). *Journal of Hellenic Studies*, 116:123–126, 1996. 22

A. Smith, B. Fuchs, and L. Isaksen. The Virtual Lightbox for Museums and Archives: a portlet solution for structured data reuse across distributed visual resources. In *Museums and the Web 2005: Proceedings*, 2005. 216

A. Smith, B. Fuchs, and L. Isaksen. VLMA: a tool for creating, annotating and sharing virtual museum collections. *Digital Medievalist*, March 21, 2008. 216

H. Smith. From farthest west. *American Journal of Archaeology*, 49(4): 463–479, 1945. 30, 230

M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering, Toronto, 2007. 114

G. Stiny. Ice-ray: A note on Chinese lattice designs. *Environment and Planning B: Planning and Design*, 4:89–98, 1977. 32

G. Stiny. *Pictorial and Formal Aspects of Shapes and Shape Grammars*. PhD thesis, University of California Los Angeles, 1975. 32

G. Stiny and W. Mitchell. The Palladian grammar. *Environment and Planning B: Planning and Design*, 5:5–18, 1978. 32

S. Tabbone, L. Wendling, and J.-P. Salmon. A new shape descriptor defined on the Radon transform. *Computer Vision and Image Understanding*, 102 (1):42–51, 2006. 115, 180

S. Tanaka, J. Kurumisawa, A. Plante, Y. Iwadate, and S. Inokuchi. Composition analyzer: computer supported composition analysis of masterpieces. In *C&C '99: Proceedings of the 3rd conference on Creativity & cognition*, pages 68–75, 1999. 213

R. Taylor, A. Micolich, and D. Jonas. Fractal expressionism. *Physics World*, 12:25–28, 1999a. 34

R. Taylor, A. Micolich, and D. Jonas. Fractal analysis of Pollock's drip paintings. *Nature*, 399:422, 1999b. 34

R. Taylor, R. Guzman, T. Martin, G. Hall, A. Micolich, D. Jonas, B. Scannell, M. Fairbanks, and C. Marlow. Authenticating Pollock paintings using fractal geometry. *Pattern Recognition Letters*, 28:269–702, 2007. 34

J. Tebbens and P. Schlesinger. Improving implementation of linear discriminants analysis for the high dimension/small sample size problem. *Computational Statistics and Data Analysis*, 52:423–437, 2007. 73

W. Technau. *Exekias*. H Keller, Leipzig, 1936. 41, 141, 162, 223

E. Tulvig. *Episodic and Semantic Memory*, pages 381–403. Academic Press, 1972. 91

M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991. 63

C. W. Tyler. Is art lawful. *Science*, 285:673–674, 1999. 197

M. van Dantzig. *Pictology: an analytic method for attribution and evaluation of pictures.* EJ Brill, Leiden, 1973. 9

M. Vickers. *Les vases peintes: image ou mirage?* Université de Rouen, 1983. 19, 21

M. Vickers. Golden Greece: relative values, minae and temple inventrories. *American Journal of Archaeology*, 94:613–625, 1990. 19

M. Vickers. Artful crafts: the influence of metalware on Athenian painted pottery. *Journal of Hellenic Studies*, 105:108–128, 1985. 19, 21

M. Vickers and D. Gill. *Artful Crafts.* Clarendon Press, Oxford, 1994. 19, 20, 21

D. von Bothmer. *Amazons in Greek Art.* Oxford: Clarendon Press, 1957. 2, 12

D. von Bothmer. review of Böhr[1982]. *American Journal of Archaeology*, 88(1):81–84, 1984. 38

D. von Bothmer. *The Amasis Painter and his world.* The J.Paul Getty Museum, Malibu, 1985. 171, 223

D. von Bothmer. A panathanaic amphora. *the Metropolitan Museum of Art Bulletin*, ns 12(2):52–56, 1954. 31

D. von Bothmer. The struggle for the tripod. In *Festchrift für Frank Brommer*, pages 51–63, 1977. 24, 236

J. Voss. Tagging, folksonomy & co-renaissance of manual indexing? In *10th International Symposium on Information Sciences, Cologne*, 2007. 217

S. Watanabi. *Knowing and Guessing:a quantitative study of inference and information.* Wiley, New York, 1969. 65, 112

A. Webb, editor. *Statistical Pattern Recognition: 2nd edition.* John Wiley and Sons, Chichester, West Midlands, 2002. 63, 64

T. Webster. *Potter and Patron in Classical Athens.* Methuen, London, 1972. 232, 235

P. Westfall, W. Johnson, and J. Utts. A Bayesian perspective on the Bonferroni adjustment. *Biometrika*, 84(2):419–427, 1997. 155

I. Widjaja, W. K. Leow, and F. cheng Wu. Identifying painters from color profiles of skin patches in painting images. Technical report, Department of Computer Science, National University of Singapore, 2003. 34

F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics*, 1: 80–83, 1945. 107

A. Willis and D. Cooper. Bayesian assembly of 3D axially symmetric shapes from fragments. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, volume 1, pages 82–89, 2004. 36

A. Willis and D. Cooper. Bayesian pot-assembly from fragments as problems in perceptual-grouping and geometric-learning. In *Proceedings of International Conference on Pattern Recognition, volume III*, pages 197–302, 2002. 36

D. Wolpert. The supervised learning no free lunch theorems. Technical report, NASA Ames Research Centre, 2001. 67

D. H. Wolpert and W. G. Macready. No free lunch theorems for search. Technical Report SFI-TR-95-02-010, The Santa Fe Institute, 1995. 67

D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1:67–82, 1997. 67

M. Yelizaveta, C. Tat-Seng, and J. Ramesh. Transductive inference using multiple experts for brushwork annotation in paintings domain. In *MULTIMEDIA '06: Proceedings of the 14th Annual ACM International Conference on Multimedia*, pages 157–160, New York, NY, USA, 2006a. 206, 208

M. Yelizaveta, C. Tet-Seng, and J. Ramesh. Semi-supervised annotation of brushwork in paintings domain using serial combinations of multiple experts. In *MULTIMEDIA '06: Proceedings of the 14th Annual ACM International Conference on Multimedia*, pages 529–538, 2006b. 206, 208, 209

H. Zhu and R. Rohwer. No free lunch for cross-validation. *Neural Computation*, 8:1421–1426, 1996. 67