



BIROn - Birkbeck Institutional Research Online

Enabling Open Access to Birkbeck's Research Degree output

Writing by numbers: case studies in digital art history

<https://eprints.bbk.ac.uk/id/eprint/47552/>

Version: Full Version

Citation: Bowman, Mike Charles (2022) Writing by numbers: case studies in digital art history. [Thesis] (Unpublished)

© 2020 The Author(s)

All material available through BIROn is protected by intellectual property law, including copyright law.

Any use made of the contents should comply with the relevant law.

[Deposit Guide](#)
Contact: [email](#)

WRITING BY NUMBERS: CASE STUDIES IN DIGITAL ART HISTORY

Mike Charles Bowman

Doctoral Thesis

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Birkbeck, University of London, 2021

ABSTRACT

While art history claims to be among the most interdisciplinary of fields, the use of digital resources and methods to create new knowledge or rethink its traditional questions is less developed than other disciplines. My thesis contributes to narrowing that gap and promotes a kind of art writing where the numerical stands alongside the verbal and the visual as an essential part of the interpretations being offered. The approach I use is the case study. Each draws on different kinds of textual and numerical data, and statistical methods for processing that data, to address some of the areas where digital art history is underdeveloped.

In the first I use statistical methods to analyse the structure and content of the catalogues and criticism of the nineteenth-century Paris Salon. The readings I develop show how that language was involved in practical, conceptual and institutional change in the nineteenth-century French art world. My readings shed new light on that art world, and extend or challenge existing scholarship. In the second I use linear regression to model auction sales with twelve contemporary artists. My models give an understanding of collectors' preferences with different characteristics of the artworks sold, and of some of the ways in which the contemporary art market has been changing. With traditional art historical methods, it is not possible to develop the kind of disaggregated perspective on collectors' preferences I present in this case study. For my third case study I created a dataset of the metadata for 59,000 artworks in the online collections of 35 modern and contemporary art museums. I use several techniques individually and in combination to identify trends in that metadata, to which I give art historical interpretations. These include correspondence analysis, topic modelling and parts-of-speech parsing. I set out a way of thinking about the history of modern and contemporary art in terms of ongoing concerns and interests which rise and fall in importance, and which cuts across conventional narratives of artist, period or movement.

My case studies are illustrations of the art historical value of using statistical methods to address textual and numerical data. They serve as examples for other art historians interested in developing that approach within their own work. There are also lessons for those working across the digital humanities from my work. My readings show how drawing together different data sources or statistical methods can support richer interpretations than comes from using them in isolation.

ACKNOWLEDGEMENTS

I would especially like to express my thanks to my wife, Evelyn, for her support and tolerance during the six years I have spent on this thesis.

I would also like to thank my supervisor, Fiona Candlin, for her support and advice, and Kate Retford and Martin Eve for their detailed comments on an early draft of Chapter 3.

TABLE OF CONTENTS

LIST OF CHARTS	7
LIST OF FIGURES	10
LIST OF TABLES	11
1. INTRODUCTION	
1.1 Digital Humanities and Digital Art History	13
1.2 Research Questions	15
1.3 Methodology	15
1.4 Writing this Thesis	18
1.5 Titles	21
1.6 Outline of Thesis and Main Conclusions	23
2. LITERATURE REVIEW	
2.1 Introduction	27
2.2 Digital Art History	30
2.3 Summary and Conclusions	54
3. THE NINETEENTH-CENTURY FRENCH ART WORLD	
3.1 Introduction	57
3.2 Salon Catalogue Entries	62
3.3 Salon Criticism	71
3.4 The Emergence and Adoption of Titling	73
3.5 Catalogue Entries and Genre	85
3.6 Male and Female Artists	102
3.7 Summary and Conclusions	107

4. THE AUCTION MARKET FOR CONTEMPORARY ART	
4.1 Introduction	111
4.2 The Auction Market, Artists, Works of Art, and Titles	118
4.3 Collectors' Preferences for Contemporary Art	124
4.4 Summary and Conclusions	147
5. MODERN AND CONTEMPORARY ART	
5.1 Introduction	151
5.2 Data Sources and Dataset Construction	155
5.3 Titling in Modern and Contemporary Art	163
5.4 The Circulation of Ideas within Modern and Contemporary Art	185
5.5 Summary and Conclusions	197
6. CONCLUSIONS	
6.1 Digital Art History	203
6.2 Digital Art History and Art History	205
6.3 Methodology	208
6.4 Summary	209
6.5 Looking Forward	209
APPENDIX A. STATISTICAL TESTING WITH SAMPLES	211
APPENDIX B. REGRESSION MODELLING OF AUCTION SALES DATA	
B.1 Introduction	213
B.2 Auction Sales Data	214
B.3 Regression Model Development	216
B.4 Regression Modelling Results	222

**APPENDIX C. CORRESPONDENCE ANALYSIS, TOPIC MODELLING AND
MODEL DEVELOPMENT, AND PARTS-OF-SPEECH TAGGING**

C.1	Correspondence Analysis	229
C.2	Topic Modelling and Model Development	231
C.3	Part-of-Speech Tagging	236
	REFERENCES	239

LIST OF CHARTS

Chart 3.1: Proportion of critics using the term 'titre' or its equivalents in their Salon reviews, 1780s and 1790s to 1860s and 1870s.	74
Chart 3.2. Proportion of entries used to refer to more than one object, 1790s to 1870s.	77
Chart 3.3. Proportion of main entries including a reference to the physical object or to the medium or type of work, 1790s to 1870s.	79
Chart 3.4. Proportion of main entries beginning with the indefinite article, 1790s to 1870s.	80
Chart 3.5. Proportion of all main entries for single oil paintings of length ten or more words, 1790s to 1870s.	81
Chart 3.6. Proportion of entries with supporting explicative text, 1790s to 1870s.	82
Chart 3.7. Proportion of main entries for single oil paintings of length three words or fewer, 1790s to 1870s.	83
Chart 3.8. Still-life entries whose only indications of subject matter were of flowers or fruit and other still-life entries, proportion of all still-life entries for single oil paintings, 1790s to 1870s.	90
Chart 3.9. Still-life entries including a reference to subjects other than flowers or fruit, proportion of all still-life entries for single oil paintings, 1790s to 1870s.	89
Chart 3.10. Proportion of landscape entries for single oil paintings containing a temporal or meteorological reference alone, a temporal or meteorological reference and a location, or a location alone, 1790s to 1870s.	89
Chart 3.11. Proportion of landscape entries for single oil paintings containing a reference to a location, split by France, Italy and Other, 1790s to 1870s.	97
Chart 3.12. Proportions of landscape entries giving a location in France including the term 'souvenir', and one or more of 'ruines', 'église', 'abbaye' and 'château', 1790s to 1870s.	98
Chart 3.13. Proportion of landscape entries for single oil paintings containing the term 'vue', 1790s to 1870s.	99

Chart 4.1. Relative prices for paintings executed at different ages by Alexander Calder and Asger Jorn.	141
Chart 4.2. Relative prices for paintings executed at different ages by Sam Francis, Martin Kippenberger, Joan Mitchell, A.R. Penck, Sigmar Polke and Cy Twombly.	142
Chart 4.3. Relative prices for paintings executed at different ages by Yoshimoto Nara, Albert Oehlen, Gerhard Richter, and Christopher Wool.	144
Chart 5.1. Correspondence analysis presenting the associations between the 122 most frequently occurring words and other linguistic components and the decades in which they were used.	166
Chart 5.2. The three dominant long-term topics in the 20-topic model, weights in the model for each decade, 1900s to 2000s	169
Chart 5.3. Entropy of the topic model distributions for each decade, 1900s to 2000s.	177
Chart 5.4. Type-Token ratio for titles in each decade, 1900s to 2000s.	178
Chart 5.5. Titles of length one, two or three words, 'untitled' alone, including a numeral, and other, proportion of all titles, 1900s to 2000s.	181
Chart 5.6. Use of 'untitled', on its own and as part of a longer title, proportion of all titles, 1900s to 2000s.	182
Chart 5.7 Entropy of the parts of speech distribution, bits, and titles of length ten words or more, proportion of all titles, 1900s to 2000s.	183
Chart 5.8. Subjective and objective parts of speech, proportion of all syntactic tokens, 1900s to 2000s.	184
Chart 5.9. Topic 1 proportion in the inferred model for national or regional collections, two-decade moving average, 1920s/1930s to 1990s/2000s.	187
Chart 5.10. Topic 2 proportions in the inferred model for national or regional collections, two-decade moving average, 1920s/1930s to 1990s/2000s.	187
Chart 5.11. Topic 3 proportions in the inferred topic model for national or regional collections, two-decade moving average, 1920s/1930s to 1990s/2000s.	188

Chart 5.12. Short titles by country or regional collection, proportion of all titles, two-decade moving average, 1920s/1930s to 1990s/2000s.	189
Chart 5.13. Titles including 'untitled' for country or regional collections, proportion of all titles, two-decade moving average, 1920s/1930s to 1990s/2000s.	190
Chart 5.14. Objective parts of speech, proportion of all syntactic tokens for each country or regional collection, two-decade moving average, 1920s/1930s to 1990s/2000s.	191
Chart 5.15. Subjective parts of speech, proportion of all syntactic tokens for each country or regional collection, two-decade moving average, 1920s/1930s to 1990s/2000s.	191
Chart 5.16. Topic 1, topic 2 and topic 3, proportions in the inferred topic model, female and male artists, 1900s to 2000s.	194
Chart 5.17. Short titles of length three words or fewer, proportion of all titles with assigned gender, female and male artists, 1900s to 2000s.	195
Chart 5.18. Titles including 'untitled', proportion of all titles with assigned gender, female and male artists, 1900s to 2000s.	196
Chart 5.19. Objective and subjective parts of speech, proportion of all syntactic tokens, female and male artists, 1900s to 2000s.	197
Chart B.1. Relative prices for paintings executed at different ages by Sam Francis.	227

LIST OF FIGURES

Figure 3.1. Extract from the catalogue for the Salon of 1849 (highlights added).	64
Figure 3.2. Extract from the catalogue for the Salon of 1806.	84
Figure 3.3. Extract from the catalogue for the Salon of 1879.	85
Figure 5.1. Word Cloud for the 50 most important words of topic 1.	171
Figure 5.2. Word Cloud for the 50 words of highest weight of topic 2.	173
Figure 5.3. Word Cloud for the 50 words of highest weight of topic 3.	174

LIST OF TABLES

Table 3.1. Salon catalogues sampled for my dataset.	66
Table 3.2. Breakdown of the core sample of catalogue entries by generic category.	68
Table 3.3. Proportion of entries for single oil paintings in each generic category, 1790s to 1810s, to 1850s to 1870s.	86
Table 3.4. Proportions of portrait entries for single oil paintings containing a reference to the social status of the sitter, by regime.	101
Table 3.5. Comparative statistics on language use by female and male artists.	103
Table 3.6. Comparative statistics on entries in each generic category for single oil paintings by male and female artists, 1790s/1810s to 1850s/1870s.	104
Table 4.1. Auction Sales Statistics for Artists included in my modelling.	127
Table 4.2. Results from Regression Modelling.	132
Table 5.1. Art Institutions in my dataset.	157
Table 5.2. Distribution of titles by decade, 1900s to 2000s.	161
Table 5.3. Number of artists with works created in each decade, 1900s to 2000s.	161
Table 5.4. Total number of titles for each country or region.	185
Table 5.5. Titles by female and male artists, 1900s to 2000s.	194
Table B.1. Results from Regression Modelling.	223

1. INTRODUCTION

Over the course of studying the history of art at various levels I have compiled an extensive collection of art historical books. Whether presenting a connoisseurial analysis of a painting, giving an iconological interpretation, taking a Marxist or feminist perspective, or drawing on queer or postcolonial theory, there is one thing almost all these monographs share: their descriptions, analyses, interpretations and arguments are developed and expressed verbally and visually. Art historians may set out narratives of singularity, specificity or exceptionality of particular artists, movements or periods. General arguments may be made through the implicit appeal to an underlying theory, or supported through the use of paradigmatic or metonymic examples. Any figures, charts and tables the author may refer to are typically supplementary to the main argument or interpretation, and are often relegated to appendices or footnotes. Reading and looking have been the primary way of acquiring knowledge in art history, and there is a strong literary tradition in the discipline. Indeed, the historical strength of this self-conception can be seen in the work of the philosopher and art critic David Carrier. In his *Principles of Art History Writing* he examines writing on art by over fifty art historians from Giorgio Vasari in the mid-sixteenth century to the 1990s, and presents his own interpretations of works by several artists as examples of what he considers to be some of the options available to the art historian to take art writing forward.¹ The only numbers that feature in Carrier's extensive quotations from other writers and in his own text are those used to identify dates or the year of publication, volume number and page numbers in references. It was reading *Principles of Art History Writing* that prompted me to consider whether the quantitative must always have that status within the discipline. What would art history writing look like if the numerical stood alongside the verbal and the visual, as an integral and essential part of the argument or the interpretation being offered?

1.1 Digital Humanities and Digital Art History

Principles of Art History Writing was published in 1994, at a time when only a few scholars were working with quantitative approaches to the humanities and when very few of the resources that might support such work were available digitally. Since then the 'digital humanities', as it has come to be known, and which can be defined as an area of scholarly

¹ David Carrier, *Principles of Art History Writing*, (University Park: The Pennsylvania State University Press, 1994).

activity that includes the creation of digital archives or databases, the development of tools to investigate them, and the use of those tools or other numerical and computational methods to address those data sources for humanistic enquiry, has developed rapidly. Online libraries such as the Hathi Trust now provide access to millions of books in digital format.² Literary historians have made extensive use of statistical text mining methods working with such digitised corpora to address a wide range of topics such as the thematic analysis of literary works and the linguistic distinctiveness of different literary genres.³ Scholars have produced websites acting as hubs for researchers in their discipline, sharing scholarship, providing information, and allowing users to carry out simple and complex searches of that information.⁴ These are just a few examples of the myriad and innovative ways the digital humanities has developed, and I will come on to discuss some more in my literature review. A perusal of the programme for any of the biennial conferences organised by the Alliance of Digital Humanities Organisations shows the large numbers of scholars and the wide range of disciplines in which digital humanities is being progressed.⁵

Digital art history, where digital resources and methods are utilised for art historical enquiry, has had a significant impact upon the discipline over the last two decades. Art historians work with digital resources on a daily basis. A digital image may supplement the original in a visual analysis. Many of the textual resources and other archival material essential for the work of the art historian are now available online. However, as I will discuss in detail in my literature review, the use of quantitative methods to address these resources and create new knowledge or to rethink its traditional questions is less developed in digital art history than in other areas of the digital humanities. Although extensive work has been done using mathematical or cartographic methods to investigate artistic networks, engagement in other ways with digital resources and quantitative methods has been more limited. The language used in digital texts of art history is an underused resource. Sophisticated statistical techniques such as those

² Founded in 2008, the Hathi Trust library includes over 8 million ebooks. It can be found at <https://www.hathitrust.org/>.

³ See, for instance, the work of scholars at the Stanford Literary Lab, which can be found at <https://litlab.stanford.edu/>, and the monographs Matthew Jockers, *Macroanalysis: Digital Methods and Literary History*, (Urbana: University of Illinois Press, 2013), Andrew Piper, *enumerations: Data and Literary Study*, (Chicago: University of Chicago Press, 2018), and Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change*, (Chicago: Chicago University Press, 2019).

⁴ One example is the British and Irish Furniture Makers Online project (BIFMO), which can be found at <https://bifmo.history.ac.uk/about/project>.

⁵ The website of the alliance can be found at <https://adho.org/>

used in the analysis of texts by literary historians have yet to be used to any extent by art historians. The art market plays a crucial role in the reception and distribution of art; however, art historians have done little to investigate that market using quantitative methods such as those used by cultural economists.

1.2 Research Questions

Digital art history, as I have characterised it, can be considered as methodological. It involves the creation or use of digital resources and methods to address those resources. It does not, in itself, prescribe the kinds of art historical enquiry with which the digital art historian might engage. And, as will be seen from my literature review, digital art historical study has been informed by theoretical commitments such as those deriving from empiricism, in which art historical concepts may be defined in terms which are directly measurable, feminism, or economics. It is also a scholarly field that remains at an early stage.

Thinking about digital art history in these terms lead me to conceive of my thesis as an opportunity to explore some of the ways digital art history can be carried out, and to the broad research questions that have guided my work. The first relate to the value of digital art history. What are the kinds of art historical question that can be addressed by statistical methods drawing on digital resources? Are some digital methods more suited to art historical enquiry than others?

The second group of questions look at digital art history in relation to the rest of the discipline. What new ways of seeing and new knowledge claims about art and its relationship to culture and history can digital methods bring into the discipline? Does it stand alone? Can it confirm, extend or challenge established art historical scholarship?

Digital art history can also be thought about as part of the digital humanities, which raises the question of whether there are methodological lessons for digital humanities scholars that can be drawn from my studies.

1.3 Methodology

The method I follow in this thesis is the case study. I chose this approach as an appropriate way of exploring the field of digital art history and addressing my research questions. Each of my three independent case studies considers areas where digital art history is underdeveloped. Each also looks at a different historical and institutional context. In

combination, my case studies were chosen to allow me to look at a number of different kinds of data source, and to explore the art historical value of the three main approaches to statistical modelling.⁶

For my case studies I have drawn on textual resources and numerical data. The data sources I have used include exhibition catalogues, critical writing, auction sales data and the metadata given in the online collections of modern and contemporary art museums. My analytical and interpretive approach has been to use statistical methods to identify trends and other patterns in these data and to give art historical readings to those trends, looking at them culturally and contextually. My case studies show that there is much of art historical value in using digital methods to examine the ways in which works of art have been described, evaluated and measured. These discourses have been integral to practical, aesthetic, conceptual, institutional and economic change in the art world,

I built a simple bespoke dataset for each case study, processing and structuring the data in ways suited to my specific research questions. In selecting my primary data resources and in structuring my datasets I faced a number of definitional questions, such as what to count as a contemporary artist. In this thesis I have taken an institutional approach to defining the meaning of these key terms, in the sense that the meaning arises from the way it is used in the particular institutional context.⁷ I define a 'contemporary artist', for instance, as someone whose works are sold at auctions advertised by the auction house as being of contemporary art. Data availability was also a constraint I had to work within. Some of the data I wanted was either not available in, or readily transformable to, a format suitable for inclusion in my datasets.

The first of the three kinds of statistical modelling I explore in this thesis is that of 'descriptive' modelling, where the researcher aims to summarise or represent different aspects of their data relevant to their interests. Often, this involves counting and categorising the data in various ways. In my first case study, for instance, I look at how the kinds of subject indicated by the catalogue entries for still-life paintings in my dataset changed over time, and give an

⁶ For a detailed discussion and comparison of these approaches, see Galit Schmeuli, 'To Explain or to Predict', *Statistical Sciences*, 25/3, (2010), pp. 289 - 310.

⁷ For institutional approaches to art see Arthur Danto, 'The Artworld', *The Journal of Philosophy*, 61/19, (1964), pp. 571 - 584, and George Dickie, *Art and the Aesthetic: An Institutional Analysis*, (New York: Cornell University Press, 1974).

art historical reading of those trends. With ‘explanatory’ modelling the researcher looks to relate a factor of interest to other factors that can be interpreted as determining or causing it to take on the values it has. These explanatory factors may come from an underlying theory the researcher is looking to establish or, less formally, from their understanding of the subject. In this thesis I use explanatory modelling of auction sales data as a way of understanding collectors’ preferences in the auction market and how the market has changed. The third approach I explore is ‘predictive’ modelling. In this approach the scholar’s aim is to develop a statistical model from a set of data that works well to predict new or future observations. I draw on a number of predictive text modelling techniques to identify patterns in the language used in the titles of the works included in my dataset of metadata. I consider what these patterns might tell us about the history of modern and contemporary art.

The sources and methods I utilise enable new ways of seeing and conceptualising art historical material. My datasets encompass broad temporal or geographical contexts. The one I have created for my third case study includes the metadata for over 59,000 works of art created in every decade from the 1900s to the 2000s and included in the online collections of art museums in nineteen different countries. My statistical methods allow me to have a wide field of view on those contexts, looking at the data in ways that bring out large-scale trends. These, in turn, lead to art historical interpretations that extend and cut across conventional narrative histories of particular artists, movements, periods or aesthetic positions. They also allow me to zero in on particular moments when there were sharp changes of trend of potential art historical interest. The various charts and tables I present in each case study provide a way of mapping out those patterns, showing the extent change happened in the historical and institutional contexts I have examined. My statistical methods also give an understanding of inter-relationships between the various features included in my datasets such as those of language use, time, location, genre, nationality, and gender. Conventional art historical methods simply do not allow such complex inter-relationships to be investigated. My readings also complement, extend and problematise earlier art historical accounts.

I carried out my data collection, dataset development and model building in two main ways. I implemented my datasets as simple one-table Excel spreadsheets. I also made extensive use of scripts I have written in the open-source programming language Python.⁹ Python scripts

⁹ The website of the Python Software Foundation can be found at www.python.org.

were used to download the auction sales data and the metadata from the online museum collections used in my case studies. They were also used to clean that data, for instance excluding works of art for which full data was not available, to partly automate the development process for my explanatory and predictive models, and to carry out text mining of documents. The analyses of my datasets were carried out using a mix of the functions available in Excel and through Python scripts. The statistical techniques I have used are all available as Python applications or computer programs in other languages such as Java and their use was built into my Python scripts. Not all the data collection and analysis I have carried out could be automated and there was also a substantial amount of manual processing involved.

1.4 Writing this Thesis

The literary historian Andrew Piper has contrasted traditional approaches to the humanities, where scholars have typically been ‘opaque about their [sources and] methods, ... and what is considered as evidence’, and the approach followed by those working in the digital humanities when it is ‘done well and done *openly*’.¹⁰ Working with complex data, developing statistical models, and interpreting the results is ‘... not fast. Rather, it slows us down and forces us to be more self-reflective’.¹¹ Piper also stresses that this way of working, where each step of the cognitive process has to be defined, is a form of knowledge mediation that can be ‘laid bare’ and ‘made mutually available’.¹²

Piper’s general observations reflect my own experience as I worked on my dataset and model development. His approach to digital literary history is one I have aspired to follow in my work on digital art history. A considerable amount of effort was required for me to define my datasets and to consolidate and clean the data from multiple and sometimes contradictory sources. Model development was an extended iterative process in which I experimented with the various variables and parameters defining the different kinds of model I have used. The data was also revisited as part of the model-building process before I arrived at the settled forms for my datasets and models that enable the readings presented in this thesis.

In my presentation of the case studies in this thesis I have endeavoured to be transparent on my data sources, and on the ways I have sampled, structured and processed the data for

¹⁰ Piper, (2018), p. 11. Piper’s italics.

¹¹ Piper, (2018), p. 11.

¹² Piper, (2018), p. 11.

inclusion in my datasets. I reflect on the consequences of these choices for the readings I offer. The statistical techniques I use provide rich descriptions of my data with considerable potential for the humanities scholar. However, they also present challenges in understanding how they work and in interpreting the results appropriately. Piper is one of several scholars working in the digital humanities who think of their statistical models as algorithms and consider the question of whether the scholar needs to have an understanding of how they work.¹³ However, this is misguided. An algorithm is merely a set of instructions to achieve a certain end. It may or may not be revealing of how the statistical model itself works or of what the model says about the data being modelled. In this thesis I do not focus on the algorithms but discuss the interpretive issues involved in the use of my statistical models, and the ways I have addressed those issues. For instance, with the explanatory technique of linear regression that I use in this thesis, the algorithm calculates the values of the parameters in the model and various associated statistics such as their statistical significance. The way in which the algorithm works is not relevant to the interpretation of the model.

Adopting an open approach to my work presented me with a challenge in writing this thesis and communicating to an art historical audience who may be unfamiliar with statistical methods and put off by too much technical detail. They may also be unused to a kind of reading that does not follow the linear, narrative pattern of conventional art history but where charts and tables need to be grasped in parallel with the main text. This is a challenge facing many scholars in the digital humanities and no cross-disciplinary approach has emerged as standard. Often the approach adopted reflects the author's interests and conceptions of the subject.

For Matthew Jockers, *Macroanalysis*, his study of the application of statistical methods to literary history, is 'fundamentally a book about method' and how those methods 'can help us to better understand and contextualise individual works' within large collections of text.¹⁴ To this end, *Macroanalysis* includes an introductory history of computational approaches to the study of literature. In each chapter he takes the reader through his modelling process, giving plain language descriptions of how the techniques he uses work, often coupled with simple

¹³ Piper, (2018), p. 11. See also Benjamin Schmidt, 'Do Digital Humanists Need to Understand Algorithms?', in Matthew K. Gold and Lauren F. Klein, *Debates in the Digital Humanities*, (Minneapolis: University of Minnesota Press, 2016), pp. 546 - 555.

¹⁴ Jockers, p. 32.

examples. He also makes frequent use of technical language. All of this is carried out in the main text, with further details in footnotes. Figures, diagrams and tables are an essential part of Jockers' mode of description and explanation, and are all presented in the main body of the book.

Andrew Piper has different aims to Matthew Jockers. Rather than positioning his book as part of the history of computational approaches in literary studies, his focus is on 'the relationship between technology and reading'.¹⁵ In *enumerations* Piper looks to place computational modelling of literature within the 'rich history of reading', through showing how it can participate in the 'construction of meaning'.¹⁶ Each chapter begins with Piper addressing what has been written on the questions he will use quantitative methods to examine. His techniques are introduced using plain language and he provides less detail in the main text on technical aspects and model building than Jockers. Summaries of model building and details of calculations are typically given in footnotes and in an appendix. Piper makes extensive use of figures, diagrams and tables in the main body of each chapter, although some are relegated to an appendix.

My answer to this challenge is to provide full details for the reader, but to only put what is necessary for my readings in the main text, and to use appendices for the remainder. For two of my case studies understanding the structure of the relevant dataset is crucial to the readings I give, and so, in the main text with both I provide an extended critical discussion of my data sources and of my datasets. With the various statistical techniques I have used, I have endeavoured in the main text to provide plain language descriptions of how they work and cover only those technical details needed to understand the interpretations I offer. Further technical details are given in the appendices. The appendices also set out the processes I went through in developing my explanatory and predictive models as examples the interested reader could follow.

The numbers, charts, figures and tables I present are all an essential part of the readings I give, and they are all included in the main body of the text setting out each case study. Unlike some other authors, who may present very complex charts with hundreds of data points with sometimes very little explanation of what is being represented, I have used my charts to give

¹⁵ Piper, p. ix.

¹⁶ Piper, p. x and p. ix.

the main trends from my modelling that are relevant to the research questions I address.¹⁷ This approach, of course, obscures the levels of variation around those trends. However, I have also looked at that aspect of my data, and when it is of art historical relevance I discuss it in the main text. With the complex or difficult to understand charts I have included in my case studies, I lead the reader through what is being presented and how to interpret them.

Part of ‘laying bare’ one’s way of working is also to make your data, methods and modelling results publicly available. Where not restricted by copyright, I have made my datasets available online in my Github repository.¹⁸ I have, however, refrained from making my Python scripts available. The reason is that many of the scripts were used for data downloading and data cleaning, and in most cases had to be tailored for the different data structures used by the online resources I have accessed. In addition, the scripts I wrote that may have a more general application such as those that automated my model development process, are mainly repeated calls to the relevant Python applications and other computer programs, and so are of little value in their own right.

The reader of this thesis will note that, unlike in much conventional art historical writing, I have not included any images of works of art. My decision to do this reflects the kinds of readings I set out in my case studies. They are all developed from textual or numerical analysis, and so images cannot play the role of exemplars of a general point as they may do in other art historical writing. They also do not involve the close reading of individual works of art and the ways in which, along with the work of art itself, textual or numerical data may contribute to our understanding or appreciation of it.

1.5 Titles

One common element to all my case studies is that the data on works of art I have collected and modelled includes their titles. Indeed, the scholarship on titles in the visual arts has prompted some of the research questions I have looked to address, and the case studies included in this thesis present aspects of the history of titling from the nineteenth century through to the current day. A small number of papers and monographs have been published on the history of titles in the visual arts, and titles often play a key role in the readings

¹⁷ This can be the case with the charts presented in the publications of the Stanford Literary Lab, which can be found at <https://litlab.stanford.edu/>.

¹⁸ My Github repository can be found at <https://github.com/MikeBowman-WritingbyNumbers>

developed by critics or art historians. All show that titles are not merely ways of identifying works of art but are of considerable art historical interest in their own right. Titles have mattered to those involved in the production, distribution and reception of art. They can indicate the subject matter of the work, have been a site for artistic innovation, can have an ideological effect, and influence the general viewer's or the critic's understanding and appreciation of the work. I will consider some of this scholarship in my literature review, and in the introduction to each case study I look at what it says that is of relevance to my work. It also provides points of comparison for the readings I develop.

In this thesis I look at titling in aggregate and identify and give art historical interpretations to large-scale and long-term patterns that can be seen in the use of titles in a number of institutional and historical contexts. To do this I needed a conceptual framework cutting across the use of titles by individual artists, or in particular periods or movements. And so, I have drawn on the work of scholars who have investigated the question of what a title is and how it functions. The literary theorist Gerard Genette was one of the first to provide such an account.¹⁹ More recently, the semiotician Josep Bepa Camprubi has surveyed the relevant literature.²⁰ Although the scholars he reviews differ in their methods of analysis and terminology, Besa Camprubi finds there is a 'remarkable convergence' between them over the three functions a title can be used to perform.²¹ The first function Besa Camprubi identifies is naming. Titles always have a nominative function. The second function of the title is the semantic one of saying something about the work it names and contributing to the meanings it is given. The title might direct the viewer's attention to particular people, objects or events being depicted, or indicate the theme of the work. Presenting a painting as 'untitled' is also a form of commentary. It might be used by the artist to indicate that the painting is to be experienced directly and in its own right. The term 'untitled' has also been recuperated by those who write about art and has become used as a title like any other. Titles can also function 'seductively', and may be used to attract the attention of the reader. A succinct allusive title in an exhibition catalogue of otherwise long descriptive titles of the paintings on show would

¹⁹ Gerard Genette, Bernard Crampé, (trans.), 'Structure and Function of the Title in Literature', *Critical Inquiry*, 14/4, (1988), pp. 692 - 720. A version of this paper appears as a chapter in Gerrard Genette, *Paratexts: Thresholds of Interpretation*, (Cambridge: Cambridge University Press, 1997).

²⁰ Josep Besa Camprubi, *Nouveaux actes sémiotiques N° 82: Les fonctions du titre*, (Limoges: Presses Universitaires de Limoges, 2002).

²¹ Besa Camprubi, p. 8.

stand out from those around it. I will have recourse to this three-fold characterisation of the functions of the title in each of my case studies.

1.6 Outline of Thesis and Main Conclusions

This introduction is followed by my literature review. In it I look critically at the scholarship in digital art history, considering what progress has been made in recent years and identifying the main gaps and underdeveloped areas. I also look at the work of some literary historians in which they draw on statistical natural language processing methods similar to those I use in one of my case studies. The results of my critical review of the literature are used to motivate the three case studies making up this thesis.

I then move on to present my case studies, each in a separate chapter. As I have already discussed, I set the scene at the start of each chapter by reviewing the art historical literature and motivate the specific research questions I address. In the main text of each chapter I set out the results of my analyses and my art historical interpretations of them. In the concluding section of each chapter I give a summary of the approach I have taken and the readings I present in the case study. I then relate it to the broad questions that have shaped my research. I consider how my work sits in relation to digital art history, to the discipline as a whole and to other relevant scholarship. I also reflect on the methodological lessons that can be drawn from the case study.

In Chapter 3 I set out my first case study. The historical and institutional context is the Paris Fine Art Salon in the nineteenth century. The textual resources I utilise are the official catalogues and critical reviews of these events. Drawing on various descriptive measures of the form and content of these resources I develop readings showing how the language used in them was involved in practical, conceptual and institutional change in the nineteenth century French art world. The specific questions I examine relate to the status of the entries in the catalogue as titles, to the artistic roles played by the language used in the entries in each of the main genres, and to the gendering of the Paris Salon in the nineteenth century. My readings also challenge some of the pre-conceptions other scholars have brought to their work on titles and the nineteenth-century Paris Salon.

The research question I address in Chapter 4 relate to what explanatory modelling can tell us about the preferences collectors at auction had for works of art presented at auction with different kinds of title, and for other characteristics of the work and the artist. The data I draw

on are the auction sales records for twelve contemporary artists, and the statistical method I use is linear regression. This is well-established with cultural economists as a way of addressing questions analogous to mine, and in the introduction, I survey the relevant scholarship. Linear regression allows me to look at the inter-relationships of characteristics of the artwork and of the artist, and at how they might have influenced the price achieved at auction. Looking at the circumstances of the sale gives some insights into the structure of the auction market for contemporary art and of the ways it has been changing.

The use of explanatory modelling in my second case study allows me to bring into art history a disaggregated understanding of collectors' preferences not possible using conventional methods. My work on preferences for different types of title is new to cultural economics. In other areas my work re-contextualises earlier studies, showing how collectors' preferences and market structures have changed.

In Chapter 5 I present my third case study. The question I address is that of what statistical methods applied to metadata can tell us about the history of modern and contemporary art. To investigate this question I constructed a dataset from the metadata for over 63,000 works of art, drawing on the online collections of 37 of the world's modern and contemporary art institutions. In this case study I use a range of descriptive and predictive techniques to identify patterns in the use of words in the titles of the works included in my dataset. Bringing these together I set out a reading of those patterns as giving of a long-term perspective on modern and contemporary art as the artistic interests signaled by the language used in titles came and went and were re-inflected, and as epistemic perspectives on the kinds of knowledge art can or should engender changed. The interpretive framework I develop in this case study is not limited to looking at titles in aggregate. In Chapter 5 I also use it to look at artists of different nationalities, and as a way of comparing the language used in titles by male and female artists.

The readings I present in Chapter 5 provide a synoptic perspective on modern and contemporary art consistent with canonical accounts of its history. Through cutting across the particularities of those accounts, they suggest some new ways of thinking about that history. The results of applying my analytical framework to artists of different nationalities

complements the growing body of work by scholars who have looked beyond the canon at the ways artists from different countries were involved in Modernism.²²

In each of my case studies I draw some comparisons between male and female artists. Our contemporary understanding is that gender is a more complex question than the binary of female and male.²³ However, it remains of interest to consider the question of whether there are any significant differences in the use of titles using the scheme that dominated public and self-representations of gender during the periods I have examined.

In the concluding chapter I draw together my three case studies and consider in the round the answers they give to the general questions that guided my work. I compare the conclusions I draw with the views expressed by other scholars working in digital art history and elsewhere in the digital humanities. To end this thesis I suggest various ways the resources and approaches I have developed might be taken forward within a digital art historical context.

²² See, for instance, Thomas DaCosta Kaufmann, Catherine Dossin and Béatrice Joyeux-Prunel, *Circulations in the Global History of Art*, (London: Ashgate, 2017), and Michele Greet, *Transatlantic Encounters: Latin American Artists in Paris between the Wars*, (New Haven: Yale University Press, 2018).

²³ For an introduction to contemporary gender studies see Raewyn, W. Connell and Rebecca Pearse, *Gender: In World Perspective*, (London: Wiley and Sons, 2014).

2. LITERATURE REVIEW

2.1 Introduction

In recent decades scholars across the humanities have increasingly engaged with the digital domain. Since the publication in 2004 of the Blackwell Companion to Digital Humanities, this term has been preferred as a description of the growing, varied and wide-ranging character of this field. Digital humanities scholars have worked on the development of remediated and born-digital archives or databases. For instance, digital libraries of electronic texts such as those maintained by the Hathi Trust or accessible through the ProQuest and Internet Archive portals have opened up large corpora to the linguist or the literary scholar as objects of study.¹ The RISM initiative is an international project which documents and provides access to musical scores held by music libraries, museums, private collections and other archives. Its database contains over 1 million records for musical scores and 200,000 records of printed music. Often, these data repositories or other applications provide sophisticated tools to search, extract and share information. One example is the Diogenes application, which provides an interface to the *TLG* and *PHI* databases of Greek and Latin texts. It allows the user to search those repositories for documents in which groups of words appear together or for all appearances of inflected forms of a given word.²

Digital humanities scholars have used computational techniques in the analysis, interpretation and presentation of their primary material. Often, especially in recent years, this has involved drawing on techniques developed by researchers in machine learning or by other computer scientists. In their work they have shed new light on established questions, confirming, extending or challenging existing scholarship, and have generated new questions. They have explored new kinds of knowledge for the humanities expressed through the use of numbers, charts, and diagrams. Scholars have employed novel means of coming to know and of communicating their work such as the use of interactive or animated displays.

¹ Founded in 2008, the Hathi Trust electronic library includes over 8 million ebooks. It can be found at <https://www.hathitrust.org/>. Launched in 1996, ProQuest provides access to over 1 million ebooks and can be found at <https://www.proquest.com/>. Also launched in 1996, the Internet archive provides access to internet sites and other cultural artifacts in digital form, including over 28 million ebooks and texts. The internet archive can be found at www.archive.org.

² The Diogenes web application can be found at <https://d.iogen.es/web>.

One example is the *Mapping the Republic of Letters* website developed by researchers at Stanford University. It presents interactive visualisations from the use of spatial mapping techniques and quantitative textual analysis, allowing the user to develop an understanding of how the exchange of letters created a virtual community of scholars in seventeenth-century Europe.³ Through the use of computer-generated 3D visualisations of Bronze Age sites, researchers in the Motion in Place project challenged established scholarship on patterns of occupation and domestic activity during that period.⁴ Statistical language processing techniques have given literary historians the opportunity to look at the inter-relationships of any of the features of their primary material that can be measured or categorised, tying together aspects of texts at all levels from the micro to the macro. For instance, as part of an analysis of 104,000 English-language works of fiction dated from 1780 to 2007, Ted Underwood, David Bamman and Sabrina Lee have looked at how conceptions of gender difference have changed over time through examining the words most strongly associated with male and female characters.⁵

In an influential 2013 paper that set out the conceptual parameters within which much of the subsequent scholarly debate has been conducted, the scholar Johanna Drucker surveyed the state of digital humanities activity in art history.⁶ The first phase of that activity had been concentrated on repository building. Although much remained to be done, this had made a substantial impact upon the discipline. As Drucker characterised it, 'Almost overnight, it seems, the inventories of museums, libraries, galleries, and collections have been digitized. ... we are suddenly able to avail ourselves of the great corpus of art historical, architectural, archaeological, and other cultural artifacts ...'.⁷ One major initiative has been the work by the Getty Research Institute to maintain and expand the Getty Provenance Index.⁸ An open access resource, the Index contains several databases with over 1 million records of interest

³ The *Mapping the Republic of Letters* website can be found at <http://republicofletters.stanford.edu/index.html>

⁴ See for instance, Stuart Dunn and Kirk Woolford, 'Reconfiguring experimental archaeology using 3D reconstruction', in *Proceedings of the 2012 Conference on Electronic Visualization and the Arts*, (New York: Springer, 2013), pp. 277 - 291.

⁵ Ted Underwood, David Bamman and Sabrina Lee, 'The Transformation of Gender in English Language Fiction', *Journal of Cultural Analytics*, February, (2018).

⁶ Johanna Drucker, 'Is There a "Digital" Art History?', *Visual Resources: An International Journal of Documentation*, 29:1-2, (2013), pp. 5 - 13.

⁷ Drucker, (2013), p. 7.

⁸ The Getty Provenance Index can be found at <https://www.getty.edu/resources/>

to art historians, including information from inventories, sales catalogues and dealer stock books for European art. Drucker drew a sharp distinction between the progress art historians and others have made in this area, which she termed ‘digitized’ art history, and what she called ‘digital’ art history proper where ‘digital methods change the way in which we understand the objects of our enquiry’.⁹ In contrast to other fields, Drucker argued, ‘... to date no research breakthrough has been made ...’.¹⁰

In this literature review I take Drucker’s 2013 paper as my starting point, along with other contemporary contributions to the debate on how the discipline of art history should respond to developments in the digital domain. I look critically at the contributions scholars have made since then to what Drucker termed ‘digital’ art history, and at the ongoing debates and controversies within this field. I have included studies by art historians and those in other fields who have presented the results of their work as being of art historical interest. I will then give my summary and conclusions of what can be learnt from the scholarship on digital art history. The three case studies I address in this thesis are introduced and positioned against these conclusions.

To identify the studies I have included in this review my starting points were the *Routledge Companion to Digital Humanities and Art History*, the online publications the *Artl@s Bulletin* (the bulletin of the ARTL@S research group), *Digital Humanities Quarterly*, *Digital Scholarship in the Humanities*, the *International Journal for Digital Art History*, the *Journal of Digital Humanities*, and *Visual Resources*, as well as the proceedings of the annual Digital Humanities Conference organised by the Alliance of Digital Humanities Organisations (ADHO) and the biennial Digital Humanities Congress organised by Sheffield University. I then used backward citation chaining beginning with the references given in those resources to identify earlier publications, and forward citation chaining using Google scholar to locate more recent studies.¹¹

⁹ Drucker, (2013), p. 7.

¹⁰ Drucker, (2013), p. 5.

¹¹ Kathryn Brown, (ed.), *Routledge Companion to Digital Humanities and Art History*, (London: Routledge, 2020). The online journals can be found at: <https://docs.lib.purdue.edu/artlas/>, <http://digitalhumanities.org/dhq/>, <https://academic.oup.com/dsh/>, <https://dahj.org/>, <http://journalofdigitalhumanities.org/>, <https://www.tandfonline.com/toc/gvir20/current>. The ADHO can be found at <https://adho.org>, and the Digital Humanities Conference proceedings at <https://www.dhi.ac.uk/dhc2022/>.

2.2 Digital Art History

Johanna Drucker's paper was one of several contributions to a special issue of the online journal *Visual Resources* dedicated to digital art history. In her thought piece the art historian Nuria Rodriguez Ortega largely agreed with Drucker. Art historians were engaging with digital material as part of their everyday practice, but technology was typically seen as a tool not as a means to develop new conceptions of art history and propose new analytical approaches. For Rodriguez Ortega, 'a true reinvention of art history should involve epistemological and methodological changes; new ways of approaching works of art; rereading of existing interpretive models; and an opening up to new interpretive paradigms, new narratives, and new discourses.'¹² She cautioned that 'issues of marginalisation needed to be addressed', with 'certain cultural contexts yet to be represented in any meaningful way' in digital art history.¹³ The question of data also needed to be looked at critically, not as a 'neutral space' but as one of 'cultural, political and ideological significance'.¹⁴ Data is not a given which presents itself free of any human bias or framing, but is constructed and always a product of interpretation which represents the world in a certain way. A two-fold approach was needed, to both explore epistemological and methodological questions, and to critically examine the inequalities and other issues this might create.

Diane M Zorich reported on a 2012 survey of art historians in the United States and the United Kingdom which confirmed the views expressed by Drucker and Rodriguez Ortega.¹⁵ The common view amongst those surveyed was that, unlike scholars in other fields in the humanities, art historians had been reluctant to embrace the transition to the digital environment. As of 2012, very few art historians had or were pursuing digital art history projects. The discipline was seen as conservative, and for some of those vested in traditional ways of working digital art history was seen as a threat.¹⁶ Through categorising and systematising primary resources it could, for instance, reduce the opportunities for serendipitous discovery. Others argued digital art history was a novelty that had yet to prove

¹² Nuria Rodriguez Ortega, 'Digital Art History: An Examination of Conscience', *Visual Resources: An International Journal of Documentation*, 29:1-2, (2013), pp. 129 - 133.

¹³ Rodriguez Ortega, (2013), p. 131.

¹⁴ Rodriguez Ortega, (2013), p. 132.

¹⁵ Diane M. Zorich, 'Digital Art History: Community Assessment', *Visual Resources: An International Journal of Documentation*, 29:1-2, (2013), pp. 14 - 21.

¹⁶ Zorich, (2013), p. 15.

its worth.¹⁷ Participants in the survey also pointed to institutional barriers to the uptake of digital art history.¹⁸ Reward and reputational structures did not encourage digital scholarship, which is often collaborative and whose products may be very different from traditional modes of publication. In contrast, art historians make their reputations through the results of solitary endeavours, publishing single-authored articles or monographs. Training in digital methods and access to technological resources were also limited.

In addition to looking at the history of the field, Drucker argued strongly in her survey for the potential of 'digital' art history to 'rethink the identity, purpose, use, and substance' of art history's objects of enquiry.¹⁹ She outlined several ways and subject areas in which this could happen, and, indeed, much of the subsequent work in 'digital' art history has been carried out along the lines Drucker suggested. Scholars have also explored some subject areas or methods that Drucker did not touch on in her survey.

I will now look at the main areas of activity in digital art history, beginning with those identified by Drucker and then looking at other work in the field. I have chosen studies which are examples of different ways that utilising digital resources and computational methods can contribute to art history, or highlight the challenges facing scholars working in this cross-disciplinary field. I will not go into the technical details here but will focus on presenting and criticising the main results and claims made in the work of the scholars I review. Where sophisticated quantitative methods are used, I have endeavoured to provide brief plain language explanations, but in some cases this has not been possible in the space available in a literature review. In some cases, scholars have used several approaches in the same study, and so the categorisation that follows is not to be seen as exclusive but as a convenient way of giving an overview of the field.

Spatio-temporal and Network Analysis

The exploration of networks is an important part of art history. Scholars look at questions such as the diffusion of technical innovations, practical know-how or aesthetic theories, and the role played in these processes by artists and critics, and by institutions such as studios, academies and artist groups. They investigate the inter-relationships of artists, dealers, patrons and other

¹⁷ Zorich, (2013), p. 15.

¹⁸ Zorich, (2013), p. 16.

¹⁹ Drucker, (2013), p. 9.

buyers and sellers of art. The histories of paintings may be considered in relation to social and economic systems of religion, nobility, conquest, and commerce. For Drucker, technology potentially allowed these and other 'networks ... from which [an art object] gains their value and meaning' to be rethought, 'in ways and at scales which go beyond traditional art historical methods'.²⁰

In digital art history there have been two main ways in which computational methods have been used to rethink questions around networks. Art historians have looked at these networks in spatio-temporal terms. Cartographic technologies allow the geographical distribution of networks and how they have changed over time to be presented on scalable maps. Scholars have also conceptualised art historical networks in abstracted mathematical terms as collections of nodes and edges linking those nodes. This has allowed them to draw on concepts developed in the mathematical theory of networks such as measures of how connected a network is or of the relative importance of different nodes.

As an example of this kind of digital art history, Drucker cited the 2012 paper by the art historians Pamela Fletcher and Anne Helmreich.²¹ In her contribution to the study, Fletcher looked at the question of what could be learnt from data drawn from annual guides, art journals, other periodicals and exhibition catalogues on the development of the art market in London over the years from 1850 to 1914. Published online, Fletcher's analysis included an animated geographical visualisation of galleries and exhibition spaces showing how the market developed and clustered around certain locations in London's West End. In contrast to previous scholarship, which was largely based on the case study, this highlighted the diversity of the London art market. It also prompted questions not addressed in that scholarship which can be answered by more traditional art historical methods, such as understanding the emergence of the 'Bond Street gallery' as a commercial identity, and the strategies of distinction employed by different galleries within the network.

Helmreich based her work on the stock books of the art dealer Goupil & Cie/Boussod, Valadon & Cie covering around 40,000 transactions involving their main Paris, London, Brussels, New York and The Hague branches over the years from 1854 to 1919. The information from these

²⁰ Drucker, (2013), p. 8.

²¹ Pamela Fletcher and Anne Heimrich, 'Local/Global: Mapping Nineteenth-Century London's Art Market', *Nineteenth Century Art Worldwide*, 11/3, (2012).

stock books is available in electronic format through the Getty Provenance Index. For each branch it includes the works bought or sold, the collectors, dealers or others involved in those transactions, and transactions with other branches in the company. Helmreich used network and spatio-temporal analysis to identify broad patterns in the corporate structure and how that changed over time. She compared cartographic visualisations of the corporate network in different periods. Treating each branch or other agents as nodes and commercial transactions as links, Helmreich turned to the mathematics of networks and the 'centrality betweenness' metric, which measures the importance of a node within a network in terms of how often it sits on the shortest paths between other nodes. These analyses showed the growing role of the London branch within the company. In the 1870s and 1880s it was, on the centrality betweenness measure, the most important branch. It also showed how artists found their largest markets in their home country, but that sales of works by foreign artists were also significant throughout the network.

One shortcoming of Heinrich's study is that she presents centrality betweenness as a given and does not engage in any methodological reflection on the choice made and the consequences for the analysis. In contrast, one scholar who has looked reflexively at her methods is Ljiljana Kolečnik. In a network analysis as part of a study into the New Tendencies art movement, she compares and contrasts what can be learnt from the use of different measures of centrality.²² In addition, the common interpretation of betweenness centrality, and the one offered by Heinrich, is that is a measure of the extent to which the agent represented by a node is playing the role of a broker between two other parties. Heinrich does not present any analysis of the stock book data to justify this reading rather than the links into and out of branches of Goupil & Co representing independent transactions between two parties in which Goupil & Co was a principal and not a broker. The art historian Sophia Quach McCabe supported her interpretation of the measure in this way through detailed examination of the archival material in her investigation of the social network around the German artist Hans Rottenhammer.²³

²² Ljiljana Kolečnik, 'The Transition of New Tendencies from Neo-Avant-Garde Subculture to Institutional Mainstream Culture. An Example of Network Analysis', in Ljiljana Kolečnik and Sanja Horvatinčić, (eds.), *Modern and Contemporary Artist Networks*, (Zagreb: Institute of Art History, 2018), pp. 84 - 112.

²³ Sophia Quach McCabe, 'Intermediaries and the market: Hans Rottenhammer's use of networks in the Copper Painting Market', *Arts*, 8/3, (2019), pp. 75 - 96.

The special edition of *Visual Resources* included a case study in which the art historians Béatrice Joyeux-Prunel, Catherine Dossin and Sorin Adam Matei introduced ARTL@S.²⁴ A major initiative in digital art history, ARTL@S began as a project in 2009 and the ARTL@S research group is ongoing. Its aims have been to study and visualise artistic circulations, constructing quantitative narratives of domestic and international exchanges. ARTL@S also makes its digital resources available to other researchers, including an open access database of nineteenth and twentieth century exhibition catalogues and a cartographic interface allowing users to search those catalogues in various ways and display the results on a scalable map.

Studying the catalogues in the ARTL@S database through the use of quantitative and cartographic methods, several scholars in the ARTL@S project have challenged dominant understandings of the history of Modernism. In one such study, Joyeux-Prunel looked at the geographical trajectory of exhibitions of modern art in Europe during the inter-war years and mapped the participation of artists in those exhibitions.²⁵ She argued that Paris was not as central as is typically considered to be the case. Indeed, it was only dominant in the formation of international careers in the artistic avant-garde in the years from 1934 to 1939.

In her study, Joyeux-Prunel draws on the ARTL@S database to make general claim but does not provide any assessment of how comprehensive that data source might be. The ARTL@S website is similarly lacking in transparency. It gives information on the number of exhibition catalogues, artists and artworks included in the database but not on how complete or representative they are of the modern art scenes in the cities where those exhibitions were held. Biased data sources are a key issue in digital art history. The art historian Michelle Moravec abandoned her study of the feminist artist Carolee Schneemann's connections to other women, as her primary data source, an edited version of Schneemann's correspondence, only included around one third of her extant letters.²⁶ Moravec recognised that any network constructed from that source would have misrepresented the scope of her contacts.

²⁴ ARTL@S can be found at <https://artlas.huma-num.fr/en/>

²⁵ Béatrice Joyeux-Prunel, 'Provincializing Paris, the Center-Periphery Narrative of Modern Art in the Light of Quantitative and Transnational Approaches', *ARTL@S Bulletin*, 4/1, (2015), Article 4.

²⁶ Michelle Moravec, 'Network Analysis and Feminist Artists', *ARTL@S Bulletin*, 6/3, (2017), Article 3.

Art historians have used spatio-temporal analyses in other ways as part of their work. It provides a way of visualising and understanding the histories of artists or artworks, as with the *Mapping Titian* project and website which presents geographical timelines of the ownership and display of Titian's paintings.²⁷ Scholars have used it to consider historically marginalised or overlooked communities. It was an essential tool in Joanna Gardner-Huggett's research into the history of the Chicago-based women artists' spaces ARC Gallery and Artemisia Gallery.²⁸ Mapping data on the artists who exhibited at these two galleries in the years from 1980 to 1985, Gardner-Huggett came to understand those spaces as primarily 'functioning within a Midwestern network of alternative spaces and strongly associated to the school of the Chicago Art Institute'.²⁹ As a result, the assumptions Gardner-Huggett had made in her earlier writing on the galleries that they were part of a national feminist art network were overturned.

The mathematics of networks has been used in novel ways to bring new knowledge into art history. Harm Nijboer, Judith Brouwer, and Marten Jan Bok used it to show how counterfactual thinking can contribute to art historical enquiry. In a presentation to the Digital Humanities 2019 conference, they described their project to 'unthink' the roles of Rembrandt and Rubens in the seventeenth century artistic communities of Amsterdam and Antwerp, the cities where they were based.³⁰ Drawing on biographical data on all known artists in those cities, they first constructed abstracted networks showing how embedded the two artists were in those communities. The networks were then reworked by removing the Rembrandt and Rubens nodes and all associated links, and their connectivity was re-assessed. This showed the communities were not predominantly centered on the two artists and highlighted the centrality of several other artists outside of the direct sphere of influence of Rembrandt or Rubens.

Network analysis has also been used to study the structure of industries. Matthew Lincoln has investigated the printmaking industry in the Low Countries, using the data held by the British Museum and Amsterdam's Rijksmuseum on fine art prints produced in the years from 1550

²⁷ The *Mapping Titian* project can be found at <http://www.mappingtitian.org/about>

²⁸ Joanna Gardner-Huggett, 'Extrapolating Influence: The Challenges of Mapping the History of ARC and Artemisia Galleries, Chicago (1980 - 1985)', *Historical Geography*, 45, (2017), pp. 37 - 65.

²⁹ Gardner-Huggett, (2017), p. 58.

³⁰ Harm Nijboer, Judith Brouwer, and Marten Jan Bok, 'Unthinking Rubens and Rembrandt: Counterfactual Analysis and Digital Art History', Digital Humanities 2019, Utrecht, The Netherlands, 8 - 12 July, 2019.

to 1750.³¹ Looking at the abstract networks in which each node was a designer, printer or publisher, and each link represented a commercial relation, he assessed how the degree of centralisation in Flemish and Dutch fine art printmaking changed over time. Lincoln used a mathematical measure of the degree of centralisation of a node as how many links it has to other nodes. At the network scale, the degree of centralisation is the ratio of the average centralisation measure for all nodes to that of the most central node. This allowed him to identify periods of increasing or decreasing centralisation, and those individuals or companies associated with major or sharp changes in the structure of the industry. His network analysis confirmed and extended established scholarship, and 'illuminated scales and dimensions otherwise difficult for art historians to conceptualize'.³²

Lincoln's study also exposes some of the limitations of network analysis. Representing a network as a homogeneous collection of edges and nodes might obscure the nature of the relations between actors which can be crucial for a more complete art historical understanding. On their own they do not allow questions such as the network dynamics and power relations in play between agents, and the meanings that circulated through the network, to be addressed. In Lincoln's case, his analysis identified periods of major structural change but did not in itself give an explanation of why that change happened. Conventional art historical methods were the way in which he explored this question and why particular printmakers had prominent roles at those times of structural change.

The final study I will look at in this brief survey of network and spatio-temporal analysis is innovative in its scope and in the new areas of art historical enquiry it opens up. However, it is one in which the scholars have not reflected adequately on the consequences of the assumptions made in structuring or conceptualising their data. In an article and associated video published in 2014, Maximilian Schich, Chaoming Song, Yong-Yeol Ang, Alexander Mirsky, Mauro Martino, Albert-László Barabási, and Dirk Helbing set out a 'Network Framework of Cultural History'.³³ Through a cartographic presentation linking the places of birth and death of more than 150,000 notable individuals over the years from 0 to 2011 CE

³¹ Matthew Lincoln, 'Social Network Centralization on Print Production in the Low Countries, 1550 - 1750', *International Journal for Digital Art History*, Issue 2, (2016a).

³² Lincoln, (2016a), p. 12.

³³ Maximilian Schich, Chaoming Song, Yong-Yeol Ang, Alexander Mirsky, Mauro Martino, Albert-László Barabási, and Dirk Helbing, 'A network framework for cultural history', *Science*, 345, (2014), pp. 558 - 562.

they set out a narrative of the cultural history of Europe and North America. However, the assumption that the history of culture can be derived from named individuals is to ignore the cultural histories of Europe and North America associated with peoples and individuals whose names are unknown to us. The early history of European culture they present is predominantly Graeco-Roman. North American cultural history only begins in 1632, and what they present as the spread of culture through that continent is silent on the impact of Western expansion on native American cultures.

In summary, the use of network analysis and spatio-temporal mapping can give new ways of understanding the often complex relationships between objects, people, institutions and events in the art world. They can make large or complex datasets visible, providing evidence for general claims, but also allowing group or individual trajectories to be traced out and seen in a broader context. Through looking at scale or through drawing on the mathematics of networks, these approaches can give answers to questions beyond the scope of traditional art historical enquiry, and can suggest questions best answered through conventional art historical methods. The examples I have reviewed also illustrate the need for digital art historians to think critically about their data sources, assumptions and methods. As will be seen, these are issues facing scholars working in all areas of the field.

Text Mining

Textual materials of all kinds play important roles in the production, display and reception of art. Resources such as art criticism, art history, philosophical writings on aesthetics, artist writings and exhibition catalogues are essential to the work of the art historian. In her survey article Drucker expressed the belief that as more of them became available online, digital engagement with the textual discourses of art history would escalate dramatically. Simply tracing changes in terminology could 'expose aspects of the field that could only be partially glimpsed through traditional reading and study'.³⁴ She also remarked that sophisticated text-mining techniques had the potential to be 'touchstones of new practice and thought'.³⁵

Much progress has, and continues, to be made in digitizing the textual resources of art history. However, digital art historical engagement with these resources in the ways that Drucker

³⁴ Drucker, (2103), p. 8.

³⁵ Drucker, (2013), p. 10.

suggested remains underdeveloped. Often studies have been led by scholars in other fields, and art historians have had little or no involvement. In my review of this area, I will look first at an example of the use of descriptive modelling to study the text of catalogues. I will then discuss three papers using sophisticated predictive text mining techniques, which are the only examples of this approach I have been able to identify for this literature review.³⁶

Drawing on collection data for New York's Metropolitan Museum of Art (the 'Met'), and online catalogues for Academy exhibitions in the United States, the United Kingdom, and France, the curator and economic historian Diana Seave Greenwald has used quantitative methods to develop 'data-driven histories of nineteenth-century art'.³⁷ Greenwald's monograph is the first to be published that takes a quantitative approach to art history. I will discuss her work on the American and British art worlds in this section, and review her study of the French art world later in this literature review as it combines text mining with econometric methods.

Looking at the online catalogues of the annual exhibitions of the National Academy of Design exhibition and for the Metropolitan Museum of Art's (the 'Met') permanent collection of American art, Greenwald investigated women artists' engagement with the nineteenth-century American art world.³⁸ Keyword searching for popular still-life subjects showed that the genre made up a much higher proportion of works exhibited by female artists than by male artists at both the Academy and the Met. The Met data includes the type of work displayed, and the most striking gender difference is that over 30% of works in the collection by women artists are miniatures, compared with fewer than 1% with male artists. Taken together the data suggests women artists were active in genres and media that were quicker to work in, which is consistent with economic theories that attribute gender inequalities in the professions in large part to the lack of time available for women to devote to their professional careers compared to men. Greenwald also notes that these same genres and media 'have traditionally been neglected in museum collections', which accounts for the lower presence of women artists in the Met's collection than at Academy exhibitions.³⁹ Greenwald recognises that the

³⁶ Museologists have also used statistical techniques to investigate the language used in museum catalogues. For a recent example see Andrew Salway and James Baker, 'Investigating Curatorial Voice with Corpus Linguistic Techniques: the case of Dorothy George and applications in museological practice', *Museum and Society*, 18(2), (July 2020), pp. 151 - 170.

³⁷ Diana Seave Greenwald, *Painting by Numbers: Data-Driven Histories of Nineteenth-Century Art*, (Princeton: Princeton University Press, 2021).

³⁸ Greenwald, (2021), pp. 85 - 114.

³⁹ Greenwald, (2021), p. 98.

conclusions she draws would be strengthened by a more complete analysis of the data sources, but that this would require a considerable amount of manual processing.

Greenwald also uses quantitative methods to investigate depictions of the British Empire at the exhibitions of the British Royal Academy from 1769 to 1910.⁴⁰ Text mining the online catalogues for those exhibitions she is surprised to find that imperial locations consistently feature in fewer than 1% of catalogue entries for works displayed, and are far less common than locations in the United Kingdom. Greenwald explores several reasons for this erasure or omission of empire. She asserts that the incompatibility between the increasingly 'inclusive' institutions of the metropole and 'extractive' institutions in the colonies made explicit images of empire inappropriate subject matter for the Royal Academy.⁴¹ In addition, practical and political constraints made it costly and, in some cases, impossible for artists to travel to imperial locations. Looking at the honorifics given in catalogue entries for portraits, she also sees the influence of inclusive institutions in the growing number of portraits that were of newly rich or ennobled people, and that of the extractive institution of the monarchy in the paucity of depictions of the Royal Family. As with her analysis of female artists in her United States, Greenwald recognises that her work is a first step in a more thorough quantitative analysis of depictions of Empire in British visual culture.

At the 2016 Digital Humanities conference the results of a collaborative project involving Artificial Intelligence researchers, museum staff and digital historians were presented.⁴² It involved the use of sophisticated text-mining methods to examine the digital archive of the Stedelijk Museum Amsterdam covering the years from 1930 to 1980s. The names appearing in these documents were extracted using a name recognition algorithm and then grouped together using a clustering algorithm, to reveal different communities of artists. The documents in the archive were also topic modelled, a statistical technique which produces a description of how words are distributed across a collection of texts. Each 'topic' is a mix of the words

⁴⁰ Greenwald, (2021), p. 115 - 152.

⁴¹ Inclusive institutions were defined by the economist Daron Acemoglu and the political scientist James Robinson as those that 'allow and encourage participation by the great mass of people ...' and extractive institutions are 'designed to extract incomes and wealth from one subset of society to benefit a different subset'. Daron Acemoglu and James Robinson, *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*, (New York: Crown, 2012), pp. 74 - 75 and 76, cited in Greenwald, (2021), pp. 124 and 125.

⁴² Smeets, J., Scholtes, J., Rasterhoff, C., and Schavemaker, M., 'SMTP: Stedelijk Museum Text Mining Project', Digital Humanities 2016, Kraków, Poland, 11 - 16 July, 2016.

appearing in the collection, and each text is mix of the topics. Looking at the most important words in the topics and the most prominent topics in each text allows the researcher to identify clusters of words that appear together across a number of documents, or are distinctive of particular texts. Plotting the important words in the topics derived from the Stedelijck's archive by time shows how the subjects and artists that were the focus of the museum's attention in exhibitions changed over the course of those fifty years. However, Smeets, Scholtes, Rasterhoff, and Schavemaker, do not go beyond the presentation of the results of their use of various text mining techniques. They do not think critically about what their results might say about the collection and exhibition policies of the museums they have looked at.

The bibliographers Carlos Garcia-Zorita and Ana R Pacios have used topic modelling to identify the key features of Mudejar art studied by art historians.⁴³ Drawing on the bibliographic records of 2,454 art historical papers on Mudejar art published in Spain between 1857 and 2013, the titles of those studies were topic modelled. The objects of study identified this way included Seville's Alcazar, religious architecture, ceramic decoration, regional variations and the Muslim tradition. In work presented to the 2017 Digital Humanities conference, the information scientists Dominic Forest, Vinh Truong, and Yvon Lemay used topic modelling to examine the text of the catalogues from exhibitions of contemporary art held at 70 artist-run centres in the Canadian province of Quebec over the years from 2000 to 2016.⁴⁴ The topics identified showed the relative importance of different media such as painting, sculpture and sound art, and of different themes including temporality, family and light, and how those changed over time.

Both Garcia-Zoriat and Pacios, and Forest, Truong and Lemay treat topic modelling as a 'black box' mechanism for identifying simple themes and trends in some textual material. Neither, like Smeets, Scholtes, Rasterhoff, and Schavemaker, do they consider the art historical value of their work. Garcia-Zoriat and Pacios do not explore what their study might say about how conceptions of Mudejar art have changed over the period included in their analysis. Forest,

⁴³ Carlos Garcia-Zorita and Ana R Pacios, 'Topic modelling characterization of Mudejar art based on document titles', *Digital Scholarship in the Humanities*, 33/3, (2018), pp. 529 - 539.

⁴⁴ Dominic Forest, Vinh Truong, and Yvon Lemay, 'De quoi est-il question dans le discours en art contemporain? La fouille de textes appliquée à l'art contemporain dans les centres d'artistes.', *Digital Humanities 2017*, Montreal, Canada, August 8 -11, 2017.

Truong and Lemay do not consider how the themes they identify compare to the views of other scholars on trends in contemporary art.

Literary historians, in contrast, have engaged in detailed and critical examinations of the working of topic modelling algorithms and of their interpretive potential for the humanities scholar. None of the art historical studies I have reviewed draw on this literature. Although not of direct relevance to digital art history, I will now look briefly at two studies in this field as examples of how sophisticated text mining techniques such as topic modelling can be used productively for humanistic scholarship. It is not my aim to provide a full survey as that would be outside of the scope of this literature review. Neither have I taken a critical stance, although there are some criticisms I would make of them in relation to the issues I have already raised with digital art history.

As with the art historical studies I have reviewed, a common use of topic modelling is for thematic analysis. Following this approach, the scholar Matthew Jockers used topic modelling to investigate the Stanford Literary Lab's digital corpus of 3,346 nineteenth-century novels written by American, British and Irish authors.⁴⁵ Restricting these novels to their nouns only, Jockers experimented with the various parameters determining how the model is constructed until a set of coherent, distinct and interpretable topics emerged, in that their most important words can be read as coalescing around themes.

In his discussion of the model results, Jockers first identified several themes differing in importance between male and female authors, and between authors of different nationalities. Female authors devote more space in their novels than men to themes such as 'Expressions of Strong Emotion', whereas male authors focus more than women on themes such as 'Villains and Traitors' and 'Enemies' (Jockers' topic names are in quotes). American authors devote much more space than British or Irish authors to the themes of 'American Slavery' and 'Native Americans', whereas Irish authors focus more on 'Tenants and Landlords'. Jocker also explored how the theme of 'Tenants and Landlords' changed in importance over time and by nationality of author. Most notably, there were several periods during the nineteenth century

⁴⁵ Matthew Jockers, *Macroanalysis: Digital Methods and Literary History*, (Urbana: University of Illinois Press, 2013) pp. 118 - 153.

when Irish novelists focused much more on this theme than at other times, with the largest peak coinciding with the Irish famine of the 1840s.

As Jockers himself observes, the results that come from looking at a small number of topics in isolation largely conform to what one might expect. However, he also looked innovatively at the large scale across the corpus and at the model as a whole, considering what it can tell us about the distinctiveness of each group of authors. To do this he fed the topic model results into a 'classifier', a machine learning technique which looks to predict the class of something, in this case the gender or nationality of the author, on the basis of various other features, in this case the prominence of each topic in a novel by the author as determined by the model. He finds that the classifier model challenges some intuitions and confirms others. American authors are the most distinguishable nationality of writer. The classifier indicated that the novels written by American authors are typified by the comparatively high presence of themes about the natural world alongside other topics. American authors, who some scholars have argued looked to emulate English writers, were far away from them in their choices of theme. Rather, Jockers argues his results give one way of thinking about how American writers in the nineteenth century responded to calls for a uniquely national literature.

In a study which goes beyond the use of topic modelling for thematic analysis, Andrew Goldstone and Ted Underwood investigated the interpretational issues involved in their use. In their study, they used the technique to investigate patterns of word use in over 21,000 articles of literary scholarship published from 1890 to 2010.⁴⁶ In developing the model they discuss in their paper, they considered the sensitivities around different choices of data source and the values of the various modelling parameters the researcher can set. Models with different numbers of topics, for instance, reveal different patterns of language use. They also considered the consequences of the simplifying assumptions made in the topic model in which texts are treated as collections of words, with no attention given to the order in which they appear in the original documents.

They stress that a topic model is not a "black box", whose pronouncements are taken on faith'.⁴⁷ Whilst they can identify patterns in word use, topic models require interpretation to

⁴⁶ Andrew Goldstone and Ted Underwood, 'The Quiet Transformation of Literary Studies: What Thirteen Thousand Scholars Could Tell Us', *New Literary History*, 45/3, (2014), pp. 359 - 384.

⁴⁷ Goldstone and Underwood, p. 366.

relate them to questions of interest to humanities scholars. For some topics, the meaning may be overt, for others covert, and other topics may be semantically broad and hard to interpret. In interpreting the model, attention needs to be given to the important words in a topic, and each topic needs to be seen in the context of the model as a whole. The same word may be prominent in more than one topic, each of which brings out different senses of the word. For instance, Goldstone and Underwood note that the word 'power' appears along with 'violence' and 'fear' in a topic which increases steadily in importance to 1980 and then declines. But it is also prominent in several other topics, including one where it is associated with 'own', 'text' and 'form', and which rises rapidly in importance in their model from 1980 onwards. They speculate that the rise in this latter topic may reflect the influence of Michel Foucault on literary studies, but the topic model shows that scholarly interest in power in one way or another was increasing steadily through the twentieth century. One of the benefits of topic modelling is that it allows such a 'nuanced, multi-faceted account of the history of scholarship' to be developed, which goes beyond the tendency to attribute change to particular scholars or schools.⁴⁸

Through other trends identified through their model they historicised many of the basic concepts of literary scholarship, showing how the nature of the discipline changed over the course of the twentieth century. Topics prominent in the early twentieth century suggest that descriptive modes of scholarship were the most important at that time. These topics then give way to ones associated with other scholarly interests. For instance, it is only in recent decades that topics where the word 'reading' is prominent become important in the model. The same holds true for 'interpretation' and the word 'literature' itself. They read the model as showing that over recent decades 'a gradual naturalization of the assumption that the discipline's aims are fundamentally critical and interpretive rather than descriptive'.⁴⁹

In their conclusion, Goldstone and Underwood argue that topic modelling can be a remarkably flexible technique allowing histories such as that of literary scholarship to be seen in a new light. The fact that numbers are involved does not give computational methods a purely objective character. The final form of the topic model is the result of many choices made by the researcher, and like any kind of evidence used by the literary scholar it requires interpretation.

⁴⁸ Goldstone and Underwood, p. 374.

⁴⁹ Goldstone and Underwood, p. 380.

Image Analysis

The close reading of images is a cornerstone of art history. In many cases the readings given by scholars will include considerations of the technique used in the artwork, the formal aspects, the subject matter and motifs, and the ways these elements are combined in the overall composition. They may play a key part in arguing for the kind of gaze a painting invites in a feminist or psychoanalytic reading.⁵⁰ An iconographic or semiotic reading may include detailed analysis of the subject matter and composition of a painting, and their relation to other depictions of the same scene, objects or events.⁵¹ In his *Principles of Art History* Heinrich Wölfflin characterised the transition in Western art from the sixteenth century to the seventeenth in stylistic terms.⁵² As will be seen, Wölfflin's formal and comparative approach has analogues in the ways several scholars have used the computational analysis of images. Other influential art historians including Alöis Riegl, Clement Greenberg and Clive Bell based their aesthetic theories on formal considerations.⁵³ Questions of style and its relation to the material, formal and iconographic content of paintings continue to engage art historians.⁵⁴

In her survey article Johanna Drucker reflected on the potential for image data mining to extend the ways visual analysis is used in art history. Through automated object recognition, large-scale trends in the circulation of patterns and motifs could become visible at geographical or temporal scales unavailable to the researcher using conventional art historical methods.

Identifying objects of particular types in an image and matching patterns across images are questions that have been studied extensively in computer vision. Computer scientists have

⁵⁰ See, for example, Marita Strucken and Lisa Cartwright, *Practices of looking: an introduction to visual culture*, (Oxford: Oxford University Press, 2000).

⁵¹ See, for example, Erwin Panofsky, *Studies in Iconology: Humanistic Themes in the Art of the Renaissance*, (London: Routledge, 1972), and Fernande Saint-Martin, *Semiotics of Visual Language*, (Indianapolis: Indiana University Press, 1990).

⁵² Heinrich Wölfflin, Jonathan Blower (trans.), *Principles of Art History: The Problem of the Development of Style in Early Modern Art*, (Los Angeles: Getty Research Institute Press, 2015).

⁵³ Alöis Riegl, Evelyn Kain, (trans.), *Problems of Style, Foundations for a History of Ornament*, (Princeton: Princeton University Press, 2018), Clement Greenberg, John O'Brian, (ed.), *The Collected Essays and Criticism, Vols 1 to 4*, (Chicago: Chicago University Press, 1988), Clive Bell, *Art*, (Delhi: Lector House, 2020).

⁵⁴ See, for instance, Jan Elsner, *Style*, in Robert S. Nelson and Richard Schiff, (eds.), *Critical Terms for Art History*, (Chicago: University of Chicago Press, 2003), pp. 98 - 109, and David Summer, *Style*, in Donald Preziosi, (ed.), *The Art of Art History: A Critical Anthology*, (Oxford: Oxford University Press, 2009), pp. 144 - 150.

looked to adapt or re-use these methods to identify the individual components of a painting, or to identify paintings with similar features. They have often presented their studies as explorations of techniques which could be of potential use to the art historian in mining large datasets of digital images. The methods utilised include different kinds of classifier, most commonly in recent years ‘convolutional neural networks’, a technique which has proven to be effective in computer vision problems such as object recognition.

The value to art historical enquiry of image analysis methods tracing out large-scale trends in the use of visual motifs or other patterns should become apparent once the methods have become reliable enough to identify the large majority of likely matches in a dataset. At present this remains a challenge for the developers of these methods. For instance, Elliot Crowley and Andrew Zisserman developed a system in which the user selects the type of object they want to search for in a dataset of target images of paintings.⁵⁵ Using the results from a Google or Bing image search for that object, a classifier is trained to recognise depictions of it at various scales and orientations using features derived from passing the images through a convolutional neural network. Crowley and Zisserman used the images of 210,000 paintings on the ART UK website as their target dataset. For most classes of object, their method achieved a success rate of around 50% for the 50 images in the dataset considered to be the most likely to contain a representation of an object of that class.

Other scholars have used computational image analysis to engage more directly with art historical questions. The use of computational techniques for the attribution of a painting to a particular period, style or movement or as being by a particular artist, have been popular topics.⁵⁶ A number of methods have been used when looking at individual artists, such as the fractal dimension of Jackson Pollock’s drip paintings.⁵⁷ For the task of classifying large collections of artworks convolutional neural networks have been popular with researchers. However, levels of accuracy with attribution remain around those of methods for object recognition, and so the art historical value of their use for this task remains an open question. Gjorgji Strezoski and Marcel Worring looked at the attribution of artist, type, school, genre,

⁵⁵ Elliott J. Crowley and Andrew Zisserman, ‘The Art of Detection’, in G. Hua and H. Dégou, (eds.), *ECCV Workshops, Part 1*, (Basel: Springer, 2016), pp. 721 - 737.

⁵⁶ For a survey see James Z. Baris Kandemir and Jia Li, ‘Computerized Analysis of Paintings’, in Kathryn Brown, (ed.), (2020), pp. 299 - 313.

⁵⁷ See for instance, Jim Coddington, John Elton, Daniel Rockmore, and Yang Wang. Multi-fractal analysis and authentication of Jackson Pollock paintings. *Proc. SPIE*, 6810, (2008), pp. 1-12.

period of creation and style or movement of the artworks in their OmniArt dataset.⁵⁸ For the main schools, their methods achieve an accuracy of between 30% and 60%. For the 985 artists with 100 or more artworks in the dataset it is 30% successful in identifying the creator.

Object recognition or image attribution are not the only use of image analysis methods of potential art historical value. In identifying components or classifying paintings, computational methods produce internal 'machine' representations of digital images which may be of art historical value in their own right. One approach several scholars have followed is that of 'operationalisation' in which a qualitative art historical concept is translated into a quantitative measure. With image analysis, scholars use the machine representations to derive the quantitative analogues. In operationalisation, the researcher needs to give careful consideration to the level of coincidence between the qualitative and quantitative concepts. This includes which aspects of the qualitative concept are likely to be captured in the quantitative measure and which are not, and what the quantitative concept might be measuring that would not be considered part of the qualitative concept.

In one of a series of papers engaging with art historical questions, a team of computer scientists at Rutgers University have looked at the question of influence.⁵⁹ Using the results of applying classifiers trained to look for objects, events, and shapes of certain types, images were given machine descriptions encoding their subject matter and composition. Influence was conceptualised as a relationship between paintings and was operationalised as a measure of the similarity of their machine descriptions.

Whilst of interest as an exploration of the methods of computer science to investigate digital images of paintings, this study falls short in the ways in which the authors address art historical questions. There are also issues around the methods and datasets used. Art historical influence is far more than visual similarity as operationalised by Saleh, Abe, Singh Arora and

⁵⁸ Gjorgji Strezoski and Marcel Worring, 'OmniArt: A Large-scale Artistic benchmark', *ACM Transactions on Multimedia, Computing, Communications and Applications*, 14/4, (2018), Article 88. Compiled by the authors from websites such as WikiArts and the online collections of a small number of art museums including New York's Metropolitan Museum of Art and MOMA, OmniArt includes around 2 million digital images.

⁵⁹ Babak Saleh, Kanako Abe, Ravneet Singh Arora, and Ahmed Elgammal, 'Toward automated discovery of artistic influence', *Multimedia Tools and Applications*, 75, (2014), pp. 3565 - 3591. In this study the authors experiment with several different approaches, and for ease of presentation I have restricted my summary to the best method they identify.

Elgammal. Artists may influence each other in a range of different ways, sharing ideas and ways of working as well as particular subject matter or compositions. Scholars have differed on the question of which artists have influenced each other. Indeed, the authors reviewed the art historical literature on the artists included in their dataset to identify what is called in computer science the ‘ground truth’ of expert beliefs on artistic influence that their methods look to match, but found little commonality.⁶⁰ Their method is also not a reliable way of identifying potential influence on the more limited measure utilised. The dataset used was the 1,710 images of paintings by 60 artists covering the years from 1400 to the present day in the Artchive database. It is small and the larger the dataset the more connections the method would identify. As the art historian Albert Boime has discussed, the French Academic system trained generations of artists in representing the same kinds of scene in the same kinds of way.⁶¹ Including paintings by these artists in a dataset would generate large numbers of spurious instances of potential influence. Even with their small dataset, the most prominent example of potential influence highlighted by the authors is between two paintings by the French Impressionist Frédéric Bazille and the American Norman Rockwell. Each painting has a similar composition for the depiction of a domestic scene. As an example of potential influence, it is one no art historian would take seriously. Bazille was active in the middle of the nineteenth century in France and Rockwell in the middle of the twentieth in the United States. Rockwell may have studied the French Impressionists, but it is highly improbable that he was influenced by one particular painting by Bazille when composing what was a common setting.

The digital cultural historians Javier de la Rosa and Juan Luis Suárez have used computational methods to investigate changes in the way in which human faces have been depicted in paintings, in what they describe as a quantitative approach to beauty.⁶² Drawing on theories from evolutionary psychology and neuro-aesthetics, they operationalised facial beauty as a measure of how symmetrically the eyes, nose and mouth were arranged on the face. Object recognition algorithms were used to identify male and female faces and their features in a dataset of 25,000 images of paintings from the thirteenth to the twentieth

⁶⁰ Saleh et al, (2014), p. 3572.

⁶¹ Albert Boime, *The Academy and French Painting in the Nineteenth Century*, (New Haven and London: Yale University Press, 1971).

⁶² Javier de la Rosa and Juan Luis Suárez, ‘A Quantitative Approach to Beauty. Perceived Attractiveness of Human Faces in World Painting’, *International Journal for Digital Art History*, Issue 1, (2015).

centuries. The symmetry measures were calculated from the features identified by the machine. What the authors find is that from the fifteenth to the eighteenth centuries the average symmetry of faces changed little, and was higher than in earlier and later centuries. Those middle centuries saw less variation in the ways in which features were arranged on the face than other periods. There was also very little difference in those trends between faces identified as male or female.

De la Rosa and Suárez's study is predicated on the assumption that the faces depicted in Western art have predominantly been those considered beautiful. They do not look to justify this assumption. Their explanation of changes in facial symmetry over time is inconsistent. At different points in their study they suggest it could be a result of variations in the 'forms in which facial beauty has been represented', or in the degree to which 'the representation of the human does not ... attempt to represent beauty'.⁶³ They do not recognise that their method does not allow these two factors to be distinguished. They also do not consider potential biases in their methods or data. For instance, they do not look at whether the facial recognition algorithm they use is itself programmed to look for symmetrical arrangements of features. Their dataset is unsourced, but all their images appear to be of paintings executed within the Western tradition, and so their study is not of 'World Painting'.

Unlike their earlier study in which the authors took their datasets as 'ground truth', the Rutgers computer scientists subsequently collaborated with an art historian and took a more critical stance on their methods and data.⁶⁴ In it the authors look at the question of how the results of applying convolutional neural networks to classify the style or movement of a painting relate to ideas of stylistic change. For their investigations they took the internal representations of the images produced by these classifiers which encoded the features used to discriminate between different styles. Almost all the variation between the machine representations of the images in their dataset could be accounted for by change in a small number of these features. What was striking was that although the machine was given no information on the date of paintings, visualisations of the two or three most important features show a clear temporal

⁶³ De la Rosa and Suárez, p. 8.

⁶⁴ Ahmed Elgammal, Marian Mazzone, Bingchen Liu, Diana Kim, Mohamed Elhoseiny, 'The Shape of Art History in the Eyes of the Machine', 32nd AAAI Conference on Artificial Intelligence, New Orleans, USA, February 2 - 7, 2018.

progression. As characterised through the machine representations used in the study, long-term stylistic change in Western art has been largely steady and continuous.

One of their main resources was WikiArt, a crowd-sourced repository of over 80,000 images of paintings and other works of art. The authors recognised that it is a very ‘noisy’ dataset. The WikiArt classifications, especially for style or movement, can overlap and have been interpreted in multiple ways by the compilers. For example, Cubism, Analytic Cubism and Synthetic Cubism are included as distinct styles. In such cases, the style categories such as these having significant overlaps were merged into one for use in the study. Styles where the paintings assigned to that category lacked art historical coherence were excluded. The authors also cautioned that these steps did not remove all the inconsistencies from WikiArt, and their results needed to be seen in that light.

To give an art historical interpretation of the machine representations the authors turned to Heinrich Wölfflin’s formal method. In his approach to comparing the ‘forms of seeing’ in different periods, Wölfflin identified five qualitative principles of visual analysis: linear/painterly, planar/recessional, closed form/open form, multiplicity/unity and absolute clarity/relative clarity. Wölfflin’s polarities were considered as the end points of a scale and 1,000 images were annotated by an art historian with a rating from 1 to 5 against each of them. Those ratings were then correlated with the most important features in the machine representations. One feature seemed to be capturing differences that were mainly what the art historian had annotated as being on the linear/painterly scale, and another the planar/recession distinction. This was confirmed by visual examination of the images displayed against those features. However, not all the variations in the machine representations correlated with Wölfflin’s categories, and so the stylistic changes between periods included in the machine representations are not fully interpretable in terms of Wölfflin’s pairs. Correspondingly, not all Wölfflin’s pairs had strong correlations with features of the machine representation. The authors suggest that these machine-produced patterns and their association with Wölfflin’s categories provide material for further investigation by art historians.

The examples I have chosen are typical of the sub-field of image analysis in that direct involvement by art historians has been very limited, with most studies involving teams of computer scientists. The example where the Rutgers computer scientists collaborated with an art historian shows the benefits of a collaborative cross-disciplinary approach in being both technically sophisticated and more humanistic in its critical and speculative stance.

Other Activity in Digital Art History

In addition to the three main areas of activity outlined by Johana Drucker some scholars have drawn on other sources of data or used other kinds of quantitative techniques. As with the other areas I have surveyed, the studies I will now review a small selection of studies which show how the methods of digital art history can bring new kinds of knowledge into the discipline, and address questions that are beyond the scope of traditional art historical enquiry.

At a conference on the Paris Fine Art Salon held in 2013, several papers taking a digital and computational approach were presented.⁶⁵ Working with the Archives des Musées Nationaux, the art historians Harriet Griffiths and Alister Mill compiled a Database of Salon Artists relating to all submissions to the Salon from 1827 to 1850.⁶⁶ In their conference papers Mill and Griffiths used descriptive statistics to analyse that database, counting and categorising the data. Mills looked at the demographics and behaviour of artists submitting works to the Salon during that period.⁶⁷ He finds that the age profile of artists submitting works changed little with an average age of 33 and that foreign-born artists made up about 12% of submitters. Mills, however, does little to develop these quantitative results and consider their art historical implications.

Through a quantitative analysis of the membership and voting records of the Salon jury during the July Monarchy, Griffiths argues that the jury was home to a wider range of views than has traditionally been acknowledged and that painter members were not as dominant in the decision-making process as has been claimed.⁶⁸ She was also able to correct the view held by several art historians that the jury consistently rejected a greater percentage of women artists than men during this period.

In *Painting by Numbers* Diana Seave Greenwald also looks at the Paris Salon, in her case drawing on a digitised version of a subject matter index of works shown at the Salon compiled

⁶⁵ James Kearns and Alister Mill, (eds.), *The Paris Fine Art Salon, 1791 - 1881*, (Bern: Peter Lang, 2015).

⁶⁶ The Database of Salon Artists can be found at <http://humanities-research.exeter.ac.uk/salonartists/about>

⁶⁷ Alister Mills, 'Artists at the Salon During the July Monarchy', in Kearns and Mill, (eds.), (2015), pp. 45 - 63.

⁶⁸ Harriet Griffiths, 'The Academy and the Salon Jury, 1831 - 1848', in Kearns and Mill, (eds.), (2015), pp. 64 - 86.

by the art historian Jon Whiteley. Her aim was to investigate links between depictions of French rural life and the countryside and socioeconomic change. The technique she used was the econometric method of regression analysis, which allows the scholar to explore possible explanatory relationships between two or more factors of interest. Whiteley's index allowed Greenwald to get a good approximation of the numbers of landscape and genre paintings at the Salon depicting particular départements. Regression analysis was used to relate those depictions to the département's transport links, presence of artists' colonies, the number of tourist attractions, and socio-economic structure. With Salon landscape painting she finds that both the more artists' colonies there were, and the lower transport costs from Paris, the more likely the département was to be depicted in landscape paintings. Long-standing tourist attractions had no effect, but the founding of a new tourist attraction increased the number of depictions of the département. Départements with the larger shares of workers employed in agriculture were less likely to be depicted, whereas those that had a legacy of strikes and labour activity were more likely to be included as subjects of landscape painting. With rural genre paintings, the strongest effects were the presence of artists' colonies and the addition of new tourist attractions, both of which were associated with higher levels of depiction at the Salon. Greenwald's results on the influence of artists' colonies are what might be expected. In other areas, as she suggests, they are not, but to understand these modelling results would require drawing on more qualitative evidence.

The art market plays a crucial role in the distribution and reception of art. However, only a few art historians have drawn on complex statistical techniques to analyse that market and bring new knowledge into the discipline. Regression modelling is one such technique. The art historian Léa Saint-Raymond used the technique to look at questions of institutional change in the French art world of the late nineteenth and early twentieth centuries.⁶⁹ In her study, Saint-Raymond challenged the established view first put forward by the economic historians Cynthia White and Harrison White that by the 1880s the Academic system in France had largely given way to the 'dealer-critic' system. In the former, careers were largely established and maintained through studying at the official School of Fine Arts and exhibiting at the official Salon. With the latter, dealers and independent exhibitions such as group shows were the

⁶⁹ Léa Saint-Raymond, 'Revisiting Harrison and Cynthia White's Academic vs. Dealer-Critic System', *Arts*, 8/3, (2019), pp. 96 - 113.

primary means of showing an artist's work, and the views of critics were crucial to an artist's reputation.

Saint-Raymond looked at general fine art auction sales in Paris in 1875, 1900 and 1925 and at sales by artists of their own works covering a similar period. Drawing on the catalogues of those sales and other metadata she compiled two datasets including, for each sale, the sales price and factors including whether the artwork had been shown at the official Salon or at another venue such as a dealer's gallery. Saint-Raymond then used regression analysis to estimate the influence of these factors on the sale price. For the general auction sales in 1875, showing at the Salon had a much stronger impact on price than showing at a dealer's gallery. In her models, it was still the most significant factor in 1900, but showing at a dealer was gaining in importance. By 1925, showing at the Salon had a negative influence on price, whereas showing at a dealer's gallery was the strongest positive factor. The artist sale data delivered similar results. Saint-Raymond argues these results suggest the Academic system was still the most important influence on artistic reputations into the early twentieth century, and it was not until after the First World War that its efficiency collapsed.

In her study Saint-Raymond is reflective on the shortcomings of her data and of her approach to comparing the importance of the Academic and dealer-critic systems. She acknowledges her work is suggestive rather than definitive and a more complete analysis would need to look at auctions and artist sales in other years and take account of the biographies of individual artists.

The scholars L. E. A. Braden and Thomas Teekens have used regression analysis in a study that looked at the strength of the associations between an artist's reputation and status and the price their works achieve at auction.⁷⁰ Drawing on the digital archives of exhibitions held at three major Dutch museums between 1930 and 1989, and on bibliographic data, they developed measures of each artist's long-term reputation and status. They conceptualised reputation as an individual measure of prestige and operationalised it in two ways: as the number of solo exhibitions held of the artist, and as the number of books dealing with the artist. Status was considered as a relational concept comparing one artist to others. It was operationalised and measured for each artist through the reputation of the other artists

⁷⁰ L. E. A. Braden and Thomas Teekens, 'Reputation, Network Status and the Art Market', *Arts*, 8/3, (2019), pp. 81 - 95.

involved in joint exhibitions with them, using the exhibition and book inclusion measures of reputation. Regression analysis was then used to relate the maximum price achieved at auction for each of these artists over the years from 1990 to 2018 to the measures of reputation and status, individually and in combination. In their models, status has a stronger impact upon prices than reputation, particularly for those artists with the lowest reputations. They also find that book mentions are a better measure of long-term prestige than appearances in solo exhibitions.

Braden and Teekens have an extended and critical discussion of concepts of reputation and status and how they have operationalised them in their work. However, they are less reflective concerning potential biases and other shortcomings of their data. The auction data they use is from sales in all countries, whereas the catalogue data comes from Dutch museums, and art museums often have a bias towards exhibitions including native or domestic artists. This bias may explain why, using their data, bibliographic information was found to be a better measure of long-term prestige than solo exhibitions. Again, prices achieved at auction can be very volatile, but they do not justify their use of the maximum price achieved rather than some other less sensitive measure such as the average price.

Other scholars have used information theory to support their enquiries. In a presentation to the 2016 Digital Humanities conference, Matthew Lincoln considered the question of subject-matter specialisation amongst Dutch painters and printmakers over the years from 1525 to 1675.⁷¹ His data sources for paintings were the Montias database of Dutch inventories from that period held by New York's Frick Museum, and the Old Netherlandish Paintings collection maintained by the Netherlands Institute for Art History (RKD). Both sources include an ascribed genre or subject matter for each of the paintings in their database. Looking at the distribution of works of different genres in the oeuvre of each artist, Lincoln calculated their diversity using the information theoretic measure of the Shannon Diversity Index or 'entropy' of those distributions. With both databases this confirmed the commonly recognised trend that Dutch painters increasingly specialised from the late sixteenth century onwards. A similar analysis of the Amsterdam Rijksmuseum's print collection showed that printmakers did not, on the whole, specialise. Lincoln argues this allowed them to respond quickly to the demands

⁷¹ Matthew Lincoln, 'If Paintings were Plants: Measuring Genre-Diversity in Seventeenth Century Dutch Painting and Printmaking', Digital Humanities 2016, Kraków, Poland, 11 - 16 July, 2016.

of what was a fast-changing marketplace. Using a different measure of diversity with the RKD's collection, a team of Dutch digital historians confirmed and extended Lincoln's results, finding that the degree of specialisation amongst Dutch painters changed little over the years from 1675 to 1750.⁷²

2.3 Summary and Conclusions

To return to Johana Drucker's question, to what extent has digital art history rethought the field's traditional objects of enquiry? Much progress has been made in spatio-temporal or network analysis, which is now an established part of art historical enquiry. The computational analysis of textual material or images is less developed. A small number of scholars have also looked at other data sources or methods. Indeed, the pool of studies from which I have drawn to illustrate particular approaches or as examples of my criticisms was much smaller in these areas than for spatio-temporal or network analysis. Nevertheless, scholars have shown the art historical value of using simple descriptive statistics, exploratory statistical approaches such as regression analysis and information theory. Art historical engagement has been least in studies utilising sophisticated predictive statistical methods for text mining and image analysis. In many cases these studies focus on technical aspects or do not engage critically with the art historical questions they present themselves as addressing.

The main debates and controversies in digital art history during the period I have reviewed have been largely around the issues raised in this literature review. A consistent theme amongst those scholars who have surveyed the field is the need for critical reflection on assumptions, methods and data, and on the biases, inclusions and exclusions created by a particular approach or reliance upon a particular data source.⁷³ The criticisms I have made of several of the studies included in this literature review reflect many of the concerns raised by these authors. Some art historians have remained sceptical about digital art history. Notably, Claire Bishop has raised the question of whether there is a 'basic incompatibility between the humanities and computational [approaches]'.⁷⁴ She argues that with the latter 'research and

⁷² Claartje Rasterhoff, Kaspar Beelen, Weixuan Li, and Ivan Kisjes, 'Measuring Innovation in the Art and Book Market during the Dutch Golden Age', Digital Humanities Benelux, Amsterdam, the Netherlands, 7 - 8 June 2018.

⁷³ A series of survey essays can be found in 'Part I: Histories and Critical Debates', in Kathryn Brown, (ed.), (2020), pp. 7 - 70.

⁷⁴ Claire Bishop, 'Against Digital Art History', *International Journal for Digital Art History*, Issue 3, (2018), p. 126.

knowledge are understood in terms of data and its exteriorisation in computational analyses' with the results that there is 'an avoidance of argumentation and interpretation'.⁷⁵ However, the studies cited by Bishop are predominantly those published in the *International Journal for Digital Art History* and include that by de la Rosa and Suarez I have reviewed, and one of the other papers published by the Rutgers team. These may be less evidence of incommensurability than, as I have discussed with de la Rosa and Suarez, of a lack of reflection on the processes of operationalisation adopted in the study. Bishop also does not consider studies published elsewhere which do take a more critical and humanistic approach. Other scholars, including Johana Drucker, accept that digital art history will not be transformative of the discipline as theoretical approaches were in the 1980s but that it has demonstrated its value.⁷⁶ Some consider the scale of the impact of digital art history on the discipline to be still an open question, for instance pointing to the relatively undeveloped state of image analysis technologies compared to those used for text mining. For Nuria Rodriguez Ortega the discipline may be nearing a turning point in the ways in which image processing is used as part of visual and formal analysis.⁷⁷

In the case studies I explore in this thesis I bring a range of different statistical techniques to bear upon textual and numerical resources. They extend digital art history in several ways. 'Digitally' through drawing on different data sources, techniques and approaches and focusing on sources and methods where art historical engagement to date has been limited. And 'art historically' through using those methods to shed new light on established questions, confirming or challenging previous scholarship, and addressing questions of art historical interest which cannot be answered using conventional art historical methods.

The textual resources I use in my first case study are the official catalogues and criticism of the Paris Salon in the nineteenth century. Using descriptive statistical methods to count and categorise the language used in those texts I present readings in which that language can be seen as an integral part of practical, conceptual and institutional change in the nineteenth century French art world. In my second study I draw on auction sales data for sales of artworks

⁷⁵ Bishop, (2018), p. 126 and p. 127.

⁷⁶ See, for instance, Johana Drucker, 'A Conversation on Digital Art History', in Matthew K. Gold and Lauren F. Klein, (eds.), *Debates in the Digital Humanities 2019*, (Minneapolis: University of Minnesota Press, 2019), pp. 321 - 334.

⁷⁷ Nuria Rodriguez Ortega, 'Image Processing and Computer Vision in the Field of Art History', in Kathryn Brown, (ed.), (2020). pp. 338 - 357.

by thirteen contemporary artists. Through regression analysis of that data I develop explanatory statistical models which give an understanding of collectors' preferences for different characteristics of those artworks, and of how the auction market for contemporary art has been changing. For my final case study I compiled a dataset of the titles and other metadata for over 59,000 artworks in the online collections of 35 major modern and contemporary art museums. I bring a wide range of different statistical and mathematical techniques to bear on this data both individually and in combination to investigate the question of what this metadata can tell us about the history of modern and contemporary art. These methods include descriptive count-based statistics, correspondence analysis, topic modelling, and parts of speech tagging. I also utilise the information theoretic measure of entropy to investigate diversity and complexity in titling.

3. THE NINETEENTH-CENTURY FRENCH ART WORLD

3.1 Introduction

When we visit a gallery or art museum our expectation is that the paintings on display will have titles. Indeed, the assumption that art works have titles is ubiquitous in the art world and has become institutionalised in the ways in which art works are presented, described, and interpreted. Titles appear on wall labels in museums and in exhibition catalogues, and may play a role in the reading of a painting given by a critic. This was not always the case, neither for artists nor for those writing about art. In the Renaissance artists were commissioned to deliver a work with a certain subject matter. In his *Lives* Giorgio Vasari referred to an artist's works in terms of their patronage, subject matter, and location. He did not use titles. For instance, in his life of Raphael the description of what we now call the *School of Athens* begins '... Raphael began a scene in the room of the Segnatura depicting the theologians reconciling philosophy and astrology with theology ...'.¹

Several scholars have given schematic accounts of the emergence of titling in the Western tradition. The art historian Stephen Bann links the emergence with that of the Academic system in France in the seventeenth century and in Britain in the eighteenth.² Under the Academic regime titling became fixed and standardised as artists were required to provide titles as marks of identification and as indications of genre and subject matter. For Ernst Gombrich, it was located in a post-Renaissance historical context in which titling was 'a by-product of the mobility of images' as paintings began to change hands more frequently and as 'artists began to create for a market and for collectors'.³ This created the need for all parties to be able to refer to a painting unambiguously. In his book-length study into titling, John Welchman largely goes along with Bann.⁴ The semiotician Leo H. Hoek points to the emergence of printing and engraving in the sixteenth century, and of French art criticism in the eighteenth century as also creating a need to refer to works of art without confusion.⁵

¹ Giorgio Vasari, *The Lives of the Artists*, Julia Bondarella and Peter Bondarella, (trans.), (Oxford: Oxford University Press, 1991), pp. 312 - 313.

² Stephen Bann, 'The Mythical Conception is the Name: Titles and Names in Modern and Postmodern Painting', *Word and Image*, 1/2, (1985), pp. 176 - 190.

³ Ernst Gombrich, 'Image and Word in Twentieth Century Art', *Word and Image*, 1/3, (1985), pp. 213 - 241.

⁴ John C. Welchman, *Invisible Colors: A Visual History of Titles*, (New Haven: Yale University Press, 1997), pp. 1 - 2.

⁵ Leo H. Hoek, *Titres, toiles et critique d'art, Déterminants institutionnels du discours sur l'art aux dix-neuvième siècle en France*, (Leiden: Brill, 2001), pp. 83 - 87. Hoek's book is not available in English.

In her 2015 *Picture Titles: How and Why Western Paintings Acquired Their Names* the scholar Ruth Yeazell problematises the histories given by these earlier authors.⁶ Through an examination of Dutch inventories, auction catalogues from France and England, catalogues from Academy exhibitions, and catalogues for art museums, Yeazell looks at the development of the language used to refer to paintings over the period from the sixteenth to the nineteenth centuries. She argues that the predominant use of the indefinite article and of brief standardised entries such as ‘a landscape’ or ‘the Holy Virgin’ in Dutch seventeenth-century inventory lists are signs of description and classification rather than of naming. The language used in the catalogues for auctions and Academy exhibitions in London and Paris in the seventeenth and eighteenth centuries, was also typically descriptive or classificatory. Yeazell notes it is in the record of a 1796 meeting of the committee charged with organising a temporary rehang of the Louvre, that the earliest reference to ‘the titles of paintings’ appears in any of the documents she reviewed.⁷ On this basis, Yeazell tentatively identifies the 1790s as the critical period during which our modern notion of the title of a painting first started to gain some traction in France at least.

The first question I address in this chapter is that of whether we can use text mining to substantiate Yeazell's observation and go beyond it to map out of the emergence of titling in the nineteenth-century French art world. Such a process would have been manifested over an extended period in changes to the linguistic practice of artists and others involved in the various textual discourses around art. And so, it is well-suited to the use of statistical methods as a way of looking systematically at the language used in those discourses at an aggregate level.

To answer it I focus on the Paris Salon. Running for up to eight weeks on an annual or biennial cycle, the Salon was the dominant institution in the French art world for the exhibition of new art through much of the eighteenth and nineteenth centuries. Without showing at the Salon and receiving recognition through jury prizes and official honours it was very difficult for an artist to sustain a career in Paris. It attracted substantial critical attention and huge crowds. It was a highly politicised event and a nexus for the often conflicting demands of Academic doctrine, artistic and critical ambition, and the cultural policy of the State.⁸ Each Salon was

⁶ Ruth Bernard Yeazell, *Picture Titles: How and Why Western Paintings Acquired Their Names*, (Princeton: Princeton University Press, 2015), pp. 25 - 51.

⁷ Yeazell, (2015), p. 48.

⁸ For detailed discussions of the history, politics and administration of the Salon see Harriet Griffiths, *The Jury of the Paris Fine Art Salon, 1831 to 1852*, (unpublished doctoral thesis, University of Exeter, 2013), Patricia Mainardi, *Art and Politics of the Second Empire*, (New Haven: Yale University Press, 1987), Patricia Mainardi, *The End of the Salon: Art and the State in the Early Third Republic*, (Cambridge: Cambridge University Press, 1993), Jane Mayo Roos, *Early Impressionism and the French State*, (Cambridge: Cambridge University Press, 1996), Gerard Monnier, *L'art et ses institutions en*

accompanied by an official catalogue that listed the artists and works on show. It was a popular publication with those attending the event and in their reviews critics often assumed the reader would have the catalogue at hand. Although there was an element of editorial control, artists were largely responsible for their own entries in the catalogue. In the nineteenth-century French art world critical reputations were established through Salon reviews, which often functioned as vehicles for critics to expound their aesthetic theories and to pronounce upon the current state of art in France.⁹ Salon reviews were published as articles in the popular press and cultural journals, or as separate pamphlets or books.

The Salon catalogues and Salon criticism are the primary sources for my investigation of the emergence of titling during the nineteenth century. The descriptive statistical methods I use involve counting and categorising these texts, and I read various changes in their content and structure as providing direct and indirect evidence of the emergence of titling. I will also look at how that process might be positioned in relation to wider structural change in the nineteenth-century French art world.

The art historical questions that can be addressed through descriptive analysis of the text of the Salon catalogues and Salon criticism are not limited to that of the emergence of titling. I use this approach to look at two other sets of questions. The first relates to issues of genre. The generic hierarchy was fundamental to discussions in the nineteenth-century French art world, including the practice of artists and the judgements of critics, who worked both within and against the system.¹⁰ It was institutionalised in Academic theory and teaching and in the medals, prizes and other honours awarded to artists at the Salon. The main generic categories provided the conceptual framework within which much of the aesthetic debate in the French art world was conducted.

Stephen Bann and John Welchman have claimed that entries in the Salon catalogue were of minimal artistic value.¹¹ For Bann and Welchman entries such as 'Vierge et enfant' [Virgin and Child] or 'Vue de la Seine' [View of the Seine] were, in the latter's words, 'essentially

France: de la Révolution à nos jours, (Paris: Gallimard, 1994), Clara H. Stranahan, *A History of French Painting*, (London: Sampson, Low, Marston, Searle, and Rivington, 1889), and William Hauptman, 'Juries, Protests and Counter-Exhibitions', *The Art Bulletin*, 67/1, (1985).

⁹ See Richard Wrigley, *The Origins of French Art Criticism from the Ancien Regime to the Restoration*, (Oxford: Oxford University Press, 1993) and Michael R. Orwicz, (ed.), *Art Criticism and its Institutions in Nineteenth-Century France*, (Manchester: Manchester University Press, 1994).

¹⁰ Griffiths, (2013), Mainardi, (1987) and (1993), and Mayo Roos, (1993), all have extended discussions of these issues of genre. See also: Lynne Ambrosini, 'Genre Painting under The Restoration and July Monarchy: The Critic Confronts Popular Art', *Gazette des Beaux-Arts*, 125, (1995), pp. 41 - 52, Stephen Bann, 'Questions of Genre in French Painting', *New Literary History*, 34/3, (2003), pp. 501 - 511.

¹¹ Welchman, (1997), p. 45.

redundant, ... merely tautolog[ies] that simply identified or repeated what was manifest in the image'.¹² Leo Hoek has looked at the institutional function of entries in the catalogue of the Paris Salon as instruments of consecration of the artists and the painting referenced.¹³ In this chapter I go beyond those accounts and address the question of what looking in aggregate and text mining the catalogue entries in each generic category can tell us about the roles they played in the construction of aesthetic value and in the institutionalisation of cultural categories. As Diana Seave Greenwald has stressed in her work on the Paris Salon, looking at catalogue entries rather than extant paintings allows a wider perspective on such issues than that given by the canonical works typically the subject of art historical enquiry.¹⁴ To support my interpretations of the language used in Salon catalogue entries I will draw on the results of text mining Salon criticism. My readings also problematise aspects of accounts of the nineteenth-century French art world, and my analysis suggests areas for further research that may be best addressed through the use of conventional art historical methods.

In nineteenth-century France it was much harder for women than for men to seek and succeed in a profession in the fine arts. Women were excluded from the School of Fine Arts, and only men were allowed to enter the Prix de Rome, the prestigious annual competition for young artists providing scholarships for them to study in Rome.¹⁵ The art historian Linda Nochlin has explored how contemporary attitudes towards class and gender further restricted women's opportunities in the fine arts.¹⁶ Rozsika Parker and Griselda Pollock have looked at how art world discourse became increasingly gendered and gendering during the nineteenth century.¹⁷ Art historians have continued to investigate the social, institutional and ideological constraints within which women had to operate and the consequences for their artistic practice. The final question I address in this case study is that of what we can learn from a statistical analysis of catalogue entries on questions of gender and the gendering of the Paris Salon in the nineteenth century. I look at whether, on the various descriptive measures I have used in my readings, there were any significant differences in the linguistic practice of male and female artists. Statistical methods also allow me to look at the levels of representation of women artists in the different genres of painting on display at the Salon and at the differences

¹² Welchman, (1997), p. 45.

¹³ Hoek, (2001), p. 57.

¹⁴ Greenwald, (2021), p. 48.

¹⁵ For a detailed survey of the training available to women artists see Charlotte Yeldham, *Women Artists in Nineteenth-Century France and England*, (London: Garland, 1984), Vol 1., pp. 40 - 62.

¹⁶ Linda Nochlin, 'Why Have There Been No Great Women Artists?', in Linda Nochlin, *Women, Art and Power*, (New York: Harper and Rowe, 1988), pp. 145 - 178.

¹⁷ Rozsika Parker and Griselda Pollock, *Old Mistresses: Women, Art, and Ideology*, (London: Pandora Press, 1981), pp. 35 - 44 and 99 - 110.

between the kinds of subject painted by women and men. As with my analysis of genre, my work extends and complements existing accounts of the nineteenth-century French art world.

To develop the readings of the Salon catalogue entries I set out in this chapter I created a dataset from the official catalogues. Drawing on the electronic versions of these catalogues, I took a random sample of 9,000 entries for paintings over the period from the 1790s to the 1870s, entering into the dataset a transcription of the main text of each entry and other information on its structure and content. The categorisations I used in my dataset reflected the questions I wanted to address. I also compiled a complete set of catalogues in text format to allow for text mining and identification of detailed trends in the use of particular words.

To complement my analysis of the language used by artists in their catalogue entries I assembled a sample of critical writing on the Salon in machine-processable format. This allowed me to carry out text searching and identify associations between the use of particular words by artists and critics. At the time of compiling my sample the number of such texts available online was very limited, and the 112 texts I assembled represent effectively all the critical writing I was able to locate. This is only a very small proportion of all the Salon reviews produced. As I will come on to discuss, in some cases the trends in the language used in critical writing are so pronounced that this small sample was sufficient for my needs, in others it precluded me from a more complete analysis.

The next two sections give critical discussions of the methodology I used to put together the samples of artistic and critical language and to structure my dataset. I then move on to look at the question of the emergence and adoption of titling, followed by my reading of how catalogue entries functioned in relation to genre. The last main section of this chapter is my discussion of questions of gender.

The case study set out in this chapter involves the use of the textual resources of exhibition catalogues and critical writing on those exhibitions, of a sampling-based approach to compiling my dataset, and of descriptive text mining techniques to model that data. The readings I give show how those texts were an integral part of practical, conceptual and institutional change in the nineteenth-century French art world. A summary and a discussion of the answers that this case study gives to the broad research questions set out in the Introduction to this thesis are given in the concluding section. I also reflect on the methodological lessons that can be taken from my work.

The descriptive statistical techniques I use in this case study involve drawing comparisons based on samples of catalogues entries and critical writing, which raises the question of how reliable the conclusions I draw are for the whole population. Statistical testing provides a way of doing this, and the trends and comparisons I identify and interpret in this chapter have been tested for their statistical significance. In Appendix A I give a brief introduction to the statistical tests I have used in this case study.

3.2 Salon Catalogue Entries

Before coming on to give my discussion of the Salon catalogue and the ways I have sampled and processed its contents to build my dataset, I will give some statistics on the population from which the sample has been drawn. From the 1790s to the 1810s the Salons were typically biennial, with five or six exhibitions per decade. There were only three Salons in the 1820s. From the 1830s the Salons increased in frequency with between six and ten held in each decade. As well as being held more often the Salon also increased in size through the nineteenth century. Both the number of artists exhibiting and the number of paintings on show rose substantially. At each Salon in the first decade of the nineteenth century the catalogue lists around 500 paintings on show by 250 painters on average, compared with averages of 2,000 paintings by around 1,000 artists in the 1830s. These figures rose to 3,100 paintings by around 1,700 artists on average listed in the catalogues for Salons in the 1870s.¹⁸ There were 333 painting members of the Republican Society for the Arts when it was established in 1791. An 1835 statistical annual on the fine arts lists 1,200 living French artists who had exhibited at the Salon, although this is likely to be a significant underestimate of the population of artists as the Salon catalogues for that period include artists not listed in the publication.¹⁹ The economists Cynthia White and Harrison White have estimated there were around 4,000 to 4,600 French painters alive in the 1860s with some sort of national recognition.²⁰

The official catalogue of the Salon, the 'livret' as it was known informally, provided the names and addresses for the artists and a listing and explication of the works on show.²¹ Throughout

¹⁸ These figures have been calculated from the Salon catalogues. There was considerable variation around these averages in the number of paintings exhibited each year depending upon the liberality of the jury and the rules governing the submission of works.

¹⁹ M. Guyot de Fère, *Statistique des Beaux-Arts en France*, (Paris: Doney-Dupré, 1835).

²⁰ Harrison C. White and Cynthia A. White, *Canvases and Careers: Institutional Change in the French Painting World*, (New York: J. Wiley and Sons, 1965), pp. 44 - 54.

²¹ The official title of the catalogue went through several close variants during the period I have examined. In 1836, for example, it was 'Explication des Ouvrages de Peinture, Sculpture, Architecture, Gravure et Lithographie des Artistes Vivans, exposés au Musée Royal, le 1st Mars 1836' [Explication of the Works of Painting, Sculpture, Architecture, Engraving and Lithography by Living Artists, exhibited at the Royal Museum, March 1st, 1836].

the period I have looked at the catalogues were variations on the same basic structure. The listing of the works on display was usually divided into sections on painting, sculpture, architecture, and one or two sections for engraving and lithography. The kinds of work listed in the painting section included oils on canvas, drawings, sketches, miniatures, watercolours, pastels, and works on enamel. In some years oils on canvas and other kinds of painting were split into two separate sections. In this case study I have taken the categorisation given in the catalogue and use the term 'painting' to refer to all kinds of items listed in the painting section. When I discuss a specific kind of painting that is given explicitly - so, for instance, when looking at paintings in oil on canvas I refer to 'oil paintings' or 'paintings in oil'.

Within each section of the catalogue exhibiting artists were listed alphabetically. Artists' names were often accompanied by additional information, which might include their address, place and date of birth, pupillage, or prizes and other honours awarded. Each name was followed by one or more entries for the works on display by that artist. Entries were numbered sequentially through the catalogue, with the numbering continuing from one section to the next, so that each entry had a unique identifier. Figure 3.1 gives an example page from the catalogue for the Salon of 1849, accompanied by my translation of that page.²² I will come on to explain the highlighting I have used for certain parts of each entry.

²² As this chapter involves an analysis of the language used by artists and critics, all direct quotations of those sources are given in the original French followed by my English translation.

10	PEINTURE.	10	PAINTING.
<p>médailles renfermant les principaux épisodes de ses derniers moments :</p> <p>1° L'archevêque bénissant les soldats à l'entrée de la place de la Bastille ; 2° L'archevêque parlant aux insurgés ; 3° L'archevêque blessé à mort ; 4° L'archevêque porté à Notre-Dame.</p>		<p>medallions containing the principal episodes from his last moments: 1* the archbishop blesses the soldiers at the entrance to the Place de la Bastille; 2* the archbishop talks to the insurgents; 3* the archbishop is fatally wounded; 4* the archbishop is carried to Notre-Dame.</p>	
<p>ALEXANDRE (LÉON), 32, rue Lafayette.</p>		<p>ALEXANDRE (LÉON), 32, rue Lafayette.</p>	
7	— Après le marché.	7	— After the market.
8	— Portrait de M ^{me} A...	8	— Portrait of Mrs A...
9	— Idem de M ^{me} M...	9	— Idem of Mrs M...
<p>AMIC (CLARICE), 60 bis, rue de Provence.</p>		<p>AMIC (CLARICE), 60 BIS, Rue de Provence.</p>	
10	— Portrait d'enfant.	10	— Portrait of a child.
<p>AMIEL (LOUIS), 14, rue de Chabrol.</p>		<p>AMIEL, (LOUIS), 14, rue de Chabrol.</p>	
11	— Frétilton, jument poulinière de pur sang, par Sylvio et Emelina, avec Light Foot, son poulain, par Alteruter.	11	— Frétilton, thoroughbred brood mare, by Sylvie and Emelina, with Light Foot, her foal, by Alteruter.
<p>ANDIRAN (FRÉDÉRIC D'), à Nantes; et à Paris, place du Louvre, chez M. Souty fils.</p>		<p>ANDIRAN, (FRÉDÉRIC D'), from Nantes; and in Paris, Place du Louvre, care of Mr Souty junior.</p>	
12	— Pastel.	12	— Pastel.
<p>Faute de vent, le Sully, de Nantes, dans les parages du Cap. ne doit son salut qu'à la perte de son mât d'arimon et à son grand foc.</p>		<p>For lack of wind, the Sully, from Nantes, in the vicinity of the Cape, owes its salvation solely to the loss of its mizzen mast and its large jib.</p>	
<p>ANDRÉ (JACQUES), 10, rue Chaptal, et 51, rue Pigale.</p>		<p>ANDRÉ, (JACQUES), 10, rue Chaptal, and 51, rue Pigale.</p>	
13	— Portrait en pied de M. d'O...	13	— Full-length portrait of Mr d'O...
14	— Idem de M. du B...	14	— Idem of Mr du B...
15	— La Sainte-Famille.	15	— The Holy Family.
16	— Une Fantaisie.	16	— A Fantaisie.
<p>ANFRAY (M^{lle} ÉLISA), 18-22, rue Beautreillis.</p>		<p>ANFRAY, (Miss ÉLISA), 18-22, rue Beautreillis</p>	
17	— La prière; étude.	17	— Prayer; study.

Figure 3.1. Extract from the catalogue for the Salon of 1849 (highlights added).

My analysis of the language used in Salon catalogue entries was based upon sampling as an effective way of approaching the sort of investigation I pursue in this chapter. Measurement of properties of a sample can inform on the whole picture and is sufficient for the identification of trends or other large-scale patterns. In my case, sampling was also necessary as it would have been too time-consuming for me to examine the whole population. Although the catalogues are available in text format, the lack of any text encoding in these documents, variations in the format of entries and the often low quality of the scanning meant I was unable to automate the extraction and analysis of text and the process I followed to compile my dataset was manual.²³ The period used in my analysis was selected initially to support my investigation into the emergence of titling. The start point was the 1790s, since my working hypothesis was Ruth Yeazell's observation that the modern notion of the title of a painting first started to gain some traction in that decade. It was also the decade in which, following the Revolution of 1789, the Salon was opened up to all practicing artists. Previously, under the Old Regime, the Salon had been restricted to members of the Academy. I chose the 1870s

²³ pdf and text versions of the catalogues are available through the Internet Archive, which can be found at www.archive.org.

as the end of the period on the grounds that it was the last full decade in which the Salon was in receipt of state sponsorship, and that from the 1880s onwards its importance diminished rapidly in the face of rival salons, and growth in group exhibitions and the commercial gallery system. In the event I did not have to deviate from my original choice, as it proved to be a suitable time frame to investigate not only the emergence of titling but also issues of genre.

Having decided upon the time period to be used in my analysis, I selected a random sample of 1,000 catalogue entries for paintings per decade, or 9,000 in total, assembled as a simple one table dataset using an Excel spreadsheet. I restricted my dataset to entries for paintings as these were sufficient for the questions I wanted to answer. If titling did emerge during the nineteenth century, then we should be able to see that process through the catalogue entries artists gave to their paintings. I could have extended it to entries for sculptures or other kinds of work on display at the Salon. However, different conventions were used with the entries for these items and one risk of including them in my sample is that the trends I look at in this case study would have been masked or harder to identify. To make the selection of items, three Salon exhibitions were chosen from each decade, and for each exhibition a random selection of 333 (or in one case 334) items was made from the catalogue entries for paintings based upon the unique number assigned to each entry in the catalogue. This allowed me to use ten-year slots as the basic unit of time in my analysis, which is appropriate when looking at long-term trends over the ninety years from the 1790s to the 1870s, and also to re-combine the Salons in other ways, such as by regime. I used three Salons from each decade rather than every year when one was held as striking a balance between a completely random selection and processing efficiency. It was quicker for me to transcribe a denser selection of entries taken from a smaller number of catalogues. This choice is unlikely to have significantly affected the robustness of my results. Table 3.1 lists the Salon catalogues I have sampled, and gives the number of paintings on display, an estimate for the number of artists showing paintings and the political regime in place at the time of the Salon.

Year of Salon	All Entries Under Painting	Artists exhibiting Paintings ²⁴	Regime
1877	3,554	2,395	Third Republic
1875	2,827	1,672	Third Republic
1872	1,536	1,009	Third Republic
1867	2,116	1,409	Second Empire
1864	2,487	1,624	Second Empire
1861	3,146	1,280	Second Empire
1859	3,045	925	Second Empire
1853	1,208	757	Second Empire
1852	1,280	777	Second Empire
1849	2,093	939	Second Republic
1846	2,107	1,125	July Monarchy
1843	1,387	779	July Monarchy
1838	2,141	1,115	July Monarchy
1836	1,836	910	July Monarchy
1834	1,348	856	July Monarchy
1827	1,052	670	Restoration
1824	1,761	420	Restoration
1822	1,348	490	Restoration
1819	1,206	446	Restoration
1814	970	458	Restoration ²⁵
1812	979	422	First Empire
1808	631	321	First Empire
1806	563	292	First Empire
1802	328	171	First Republic
1799	354	199	First Republic
1798	428	189	First Republic
1795	535	179	First Republic

Table 3.1. Salon catalogues sampled for my dataset.

My sample was adequate for the analyses that contribute to most of the readings I set out in this chapter. However, it included only 60 entries per decade on average for oil paintings by female artists. To allow for robust comparisons between male and female artists I boosted this to 200 for each decade from the 1800s to the 1870s. There were only 115 entries in total for oil paintings by female artists at all the Salons in the 1790s, and I entered all of these into my

²⁴ Estimated from the number of artists per page for the first 50 pages, multiplied by the total number of pages in the painting section.

²⁵ The Salon of 1814 began on 1 November during Napoleon's exile when Louis XVIII ruled France. The monarchy under Louis was re-instated in 1815 after Napoleon's final defeat at Waterloo.

dataset. In what follows I will refer to the sample of 9,000 entries as my 'core' sample or 'core' dataset. It is the sample that has been used in all sections of this chapter other than Section 3.4, where I make my comparison of male and female artists.

For each catalogue entry I recorded in my dataset the year of the Salon, the catalogue number, and a transcription of the text in the main part of each entry in the sample, that is the continuous piece of text beginning on the same line as the catalogue number. I will refer to this unit of text as the 'main entry'. Text relating to the medium or type of work was excluded unless it was the only text or was an integral part of the main entry. I chose this as the unit of text for transcription as it was the part of the catalogue entry that became the title. In Figure 3.1 I have highlighted the text constituting the main part of each entry.

One of the measures of change I look at in this chapter is the word length of main entries. For most types of entry the word content was self-evident. Phrases with apostrophes such as 'd'enfant' [of a child] in entry number 11 in Figure 3.1 were counted as two words. In the case of entries for portraits, where many followed the format of items 8, 9, 13 and 14 in Figure 3.1 and provided a gender prefix and the initials of the sitter, the prefix and each initial were counted as separate words. I did not include as a word the 'trois étoiles' [three stars] device of *** that appeared often in portrait entries such as number 13 and 14 of Figure 3.3. For example, the word count for entry 7 in Figure 3.1 is 3 and for entry 13 is 7. I applied the same rule to instances where the names of individuals appeared in other entries, such as the naval commanders in an entry for a depiction of a sea battle.

Where statistics relevant to the questions I wanted to address could not be calculated directly from the text of the main entry, I recorded a number of other fields in my dataset. I assigned a generic category to each entry. However, this was not straightforward. One complication is that the system of genres was used to classify paintings and was not a taxonomy of catalogue entries. The same catalogue entry could be used for paintings with the same subject matter that would have been regarded as works in different genres. As images and other details such as size are not available for most of the paintings referenced in the catalogue I could not, for instance, carry over 'history' as a category to apply to entries. Executed on the grand scale a depiction of a biblical or mythological scene would have been regarded as a history painting. A small-scale work with the same entry in the catalogue would have been considered a historical genre painting. In defining the categories of catalogue entry I had to adapt those used for paintings, whilst looking to stay as close to them as possible. The second issue was that of which kind of scheme to use. As my aim was to identify long-term trends and other

patterns in the language used by artists, the categories needed to be applicable to entries over the whole of the period of interest. I also needed to have sufficient entries in each category to allow for a robust trend analysis. Together these requirements indicated I should use a small number of categories. The generic categories I defined therefore reflected the basic categories of history, portrait, landscape, genre, and still-life used in Academic writing and by critics and artists.²⁶ I did not include in my scheme sub-divisions of those categories, such as distinguishing landscapes and seascapes, nor short-lived genres such as the ‘fantaisie’, which came into and out of fashion. Table 3.2 gives the breakdown of my core sample by the generic category assigned to each entry.²⁷ I will now give my definitions of each of the generic categories of entry and my discussion of how these categories relate to the main genres of painting.

Category	Landscape	Historical	Genre	Portrait	Still-life	Other	Total
Number of entries	2,714	1,614	1,500	2,308	419	445	9,000
Proportion of total	30%	18%	17%	26%	5%	5%	100%

Table 3.2. Breakdown of the core sample of catalogue entries by generic category.

The **landscape** category comprised entries describing places in a natural or built world, either real or imagined, for example ‘*fôret de Fontainebleau, près du Mont Ussy*’ [the Fontainebleau Forest, near Mount Ussy], or natural phenomena such as ‘*l’orage*’ [the storm], or giving no more than the genre of the work as in ‘*paysage*’ [landscape] or ‘*marine*’ [seascape]. Entries giving animal subjects such as ‘*vaches au pasture*’ [cows at pasture] were included in this category as well as entries giving the subject as the interior of public or religious buildings. I categorised as landscape entries such as ‘*paysage historique representant philoctete dans*

²⁶ For definitions of these terms in Academic theory see the 1792 *Dictionnaire des Arts de Peinture, Sculpture et Gravure* [Dictionary of the Arts of Paintings, Sculpture and Engraving], compiled by the Academicians Pierre-Charles L vesque and Claude-Henri Watelet. Pierre-Charles L vesque and Claude-Henri Watelet, *Dictionnaire des Arts de Peinture, Sculpture et Gravure*, (Paris: Prault, 1792), Vols 1 to 4.

²⁷ Working with the Archives des Mus es Nationaux, the art historians Harriet Griffiths and Alister Mill compiled a Dataset of Salon Artists relating to all submissions to the Salon from 1827 to 1850, drawing upon archive material including the Register of Works submitted to the jury for each Salon. From 1840 to 1850 most works submitted were assigned a genre in the Register, utilising a wider range of categories than in my analysis, and reflecting the prizes given to artists at that time. For instance, the Register classifies works as landscapes and seascapes, which are both included in my landscape category. Griffith and Mill extended the categorisation to the earlier years in their dataset. My categorisation was carried out independently of that in the Dataset of Salon Artists, however, a comparison for two years indicated a high degree of coincidence. The Database of Salon Artists can be found at <http://humanities-research.exeter.ac.uk/salonartists/about>.

l'île de Lemnos' [historical landscape representing Philocrete on the Isle of Lemnos], that indicated the painting was a historical landscape in which small religious, mythological or historical figures were presented in an idealised classical landscape. Entries giving no more than a time of day or a season such as 'printemps' [spring] were also included.

My assignment of the landscape category to an entry was also not always clear-cut as artists could produce works with entries cutting across that category and others. One example is 'un soir de vendange à Mont-Saint-Père', [the evening harvest at Mont-Saint-Père], which includes genre and landscape references. As I wanted to investigate the long-term trends in how catalogue entries functioned to signal aesthetic value within landscape painting and not into how they functioned in relation to sub-genres such as portrayals of rural labour, I assigned 'un soir de vendange, à Mont-Saint-Père' to the landscape category and not the genre category. Another example is that entries such as 'printemps' were sometimes used for genre works or allegorical paintings rather than landscapes. These choices will have introduced some biases into my analysis, for instance, with the first example I have discussed, boosting the proportion of landscape entries and cutting that of the genre category, but the proportion of entries where I had to make such a decision was small.

Entries in the **historical** category took the form of a description of the individuals and actions depicted. It included entries describing subjects taken from the bible or from classical history and myth. Allegorical, sacred, and everyday subjects from French history such as saints, monarchs and well-known historical figures were also included in this category, as were recent battle scenes, other patriotic subjects, and subjects from the history of other countries. I also categorised as historical entries those giving subjects taken from works of historical fiction such as those by William Shakespeare, Walter Scott, and Claude Bonet. The types of painting whose catalogue entries I have classified as 'historical' include history paintings, historical genre works, and historical landscapes where, as was sometimes the case, the entry was not explicit that the painting was a landscape.

I took **genre** entries to be those describing domestic, rural or urban everyday events involving human activity such as having lunch, going to market, working the land, attending church or school, or hunting. Entries not indicating the manifest subject matter of the painting, but which were humorous, sentimental, or moralizing, such as '«cet âge et sans pitié»' ['this age is without pity'], were typically used for everyday genre paintings and were all assigned to this category. I also included entries giving generic military subjects such as soldiers at camp.

I used the **portrait** category to record catalogue entries of the form ‘Portrait de/d’ ...’ [Portrait of ...]. I also assigned entries that did not begin that way but were most likely the name of a living person or referred to a specific group of individuals such as ‘tableau de famille’ [family portrait] to this category. Entries containing portrait and other genre references, as with ‘Portrait de M. J. D. dans un paysage’ [Portrait of Mr J. D. in a landscape], were classified as portraits.

Entries giving descriptions of immobile or inanimate natural or man-made objects such as ‘gibier et fruits’ [game and fruit] or ‘fleurs et vase’ [flowers and vase], or giving a type of painting with such subjects, as in ‘tableau de fruit’ [fruit painting] or ‘nature morte’, were assigned to the **still-life** category.²⁸

The remaining entries I classified as **other**. This included entries not specific as to the subject matter such as ‘étude’ [study] or ‘cadre de miniatures’ [framed miniatures] and those indicating a subject matter not included in the other categories such as paintings made ‘after’ those of renowned artists, botanical studies and portraits of household pets.

With some of the generic categories I recorded additional characteristics of the entries. For landscape entries I noted if it included a reference to the hour of day, time of year, or weather conditions as in ‘soleil couchant’ [setting sun], ‘fôret en automne’ [forest in Autumn] or ‘effet de neige’ [snow effect]. I also noted the location of the site depicted, if given, as in France, Italy, another European country, or elsewhere. With portrait entries I recorded the gender and social position, if given, of the sitter. The categories used for social position were noble, royal or imperial; military; religious; representative or administrative; professional; artistic; other; and, not given.

I included three other characteristics of each catalogue entry in my dataset. I recorded the medium or form of the work given in the entry, such as an oil painting, watercolour, miniature, study, or drawing. I noted the number of objects relating to the catalogue entry. An example of an entry for more than one object is ‘trois portraits sous le même numéro’ [three portraits under the same number].²⁹ I also recorded whether or not the catalogue entry included a

²⁸ I have left ‘nature morte’ untranslated since, as will be seen, it was used in different senses by both critics and artists and to render it as ‘still-life’, which is the standard translation, would be to favour one of those senses and exclude others.

²⁹ In this case study I have only used the split of entries referring to paintings in oil and those in other media, and the split between entries referring to one painting or to more than one in this case study. However, I have retained the more detailed categorisations in my dataset as they may be of value to other scholars.

detailed exposition of the subject matter in addition to the main entry. For instance, history paintings might be accompanied by a detailed explanation of the actions depicted or an entry for multiple objects might be followed by a list giving their subjects. Entry number 12 in Figure 3.1 is an example of an entry providing a detailed explication.

Finally, I recorded the gender, male or female, of the artist as given in the catalogue. In almost all cases, catalogue entries explicitly identified female artists through the inclusion of a gender prefix, either 'Citoyenne' [Citizen] (sometimes abbreviated to 'C^{nne}') during most years of the First Republic or 'Madame' [Mrs] and 'Mademoiselle' [Miss] (usually abbreviated to M^{me} or M^{lle}) at other times. In almost all years, entries for male artists simply provided their names.

As I have already discussed, the other personal details recorded on artists in the catalogue included, in most years, their address and, in some years, their nationality and status as medal winners or holders of the Legion of Honour. The size and the ownership of a painting, including those purchased by the state, might sometimes also be identified. These characteristics were not included in my dataset as they were not relevant to my research questions or were not present over a long enough period to enable any trends to be identified.

The electronic versions of the catalogue are all available as text documents, and this allowed me to carry out further investigations of linguistic change through searching on the occurrences of key words or phrases. Text searching enabled an approximately complete count for each catalogue and could be carried out for every year with a Salon. It was called for in some situations, such as those where there was a sharp change of trend from one year to the next, that I wanted to investigate. As I have already mentioned, the quality of the document scanning varied from year to year and text searching did not always identify every occurrence of the relevant piece of text. However, testing with a small number of catalogues indicated it did return the large majority.

3.3 Salon Criticism

Salon reviews were published as articles in the popular press and cultural journals, or as separate pamphlets or books. They took a variety of forms: some were simply annotated catalogues, others were comical or satirical verse, and many of the longer reviews would begin with general remarks from the author followed by reviews of the works considered worthy of mention, either favourably or unfavourably. The Cambridge bibliographies of Salon criticism list around 400, 500 and 900 critics active in the periods covering the Old Regime to the

Restoration, the July Monarchy to the Second Republic, and the Second Empire respectively, and around 200, 450 and 450 publications in which their reviews appeared.³⁰

The base used for my analysis of critical language was a sample of Salon reviews covering the decades from the 1780s to the 1870s.³¹ I wanted these to be in electronic format to allow me to conduct keyword searching as a complement to my analysis of catalogue entries. I wanted a range of different types of criticism, from newspaper articles to pamphlets and book-length reviews. I did not include reviews by the same critic made in neighbouring decades, and applied a similar criterion to anonymous reviews appearing in the same journal, as these may have also been by the same person. This constrained my sample to 112 documents in total.³²

I assembled the sample from a number of online libraries and archives, mainly the Bibliothèque Nationale de France's electronic library, supplemented by the Internet archive and Google books. Most of these electronic documents were in text format. However, some, especially from the earlier decades, were pdfs and not text searchable, and in these cases they were searched manually. Of the 112 documents, 94 were published as books or pamphlets, mainly contemporary publications but also some compilations of earlier Salon reviews, and 18 were articles from newspapers and journals. The critic was anonymous with 41 reviews, with the majority of these in the decades from the 1790s to 1810s.

The critics included in my sample expressed a variety of aesthetic views. They ranged from those such as Étienne Jean Délécluze and Paul Mantz who defended the traditional principles associated with the Academy, to Théophile Thoré and Jules-Antoine Castagnary who challenged those principles and championed emerging tendencies or schools such as Naturalism and Realism, as well as those who occupied the middle ground such as Gustave Planche who looked to reconcile tradition with Naturalism. This is, of course, a highly stylised characterisation of the positions adopted by critics, which was typically much more nuanced in terms of their aesthetics and their relations to the cultural politics of their period. Critics'

³⁰ Christopher Parsons and Martha Ward, (eds.), *A Bibliography of Salon Criticism in Second Empire Paris*, (Cambridge: Cambridge University Press, 1986), Neil McWilliam, (ed.), *A Bibliography of Salon Criticism in Paris from the Ancien Regime to the Restoration, 1699 - 1827*, (Cambridge: Cambridge University Press, 1991). Neil McWilliam, (ed.), *A Bibliography of Salon Criticism in Paris from the July Monarchy to the Second Republic, 1831 - 1851*, (Cambridge: Cambridge University Press, 1991). The totals exclude anonymous critics and will multiply count critics active, and periodicals published, in more than one period and critics who published under more than one name.

³¹ I included the 1780s in my analysis of critical language to allow a split into five twenty-year periods. Unlike the Salon, there was not a sharp break in critical writing following the Revolution of 1789.

³² A list of the Salon reviews included in my sample is given in the References.

positions could also change over the course of time, as could the political alignments of different aesthetic positions. Scholars have also differed over the categorisation of critics.³³ However, the arguments for and against any one particular categorisation are not directly relevant to my investigations in this chapter. The key point is that my sample of critics is not strongly biased towards those of one particular critical position.

3.4 The Emergence and Adoption of Titling

Critical language

The simplest and most direct approach to measuring and mapping out the emergence of titling is to look at the use of the noun 'titre' [title] and related words such as the verb 'intituler' [entitle] by critics in their references to paintings on display at the Salon. The results of my analysis of their use are exhibited in Chart 3.1, which looks at the 112 reviews in my sample grouped together in five twenty-year periods from the 1780s/1790s to the 1860s/1870s. For each period it presents the proportion of critics using the term 'titre' or an equivalent in their reviews of the Salons at that time. As can be seen, critical use was very rare in the 1780s and 1790s. Of the 19 examples of critical reviews from the 1780s and 1790s in my sample, only one, from 1793, referred to the 'titres' of art works on display at the Salon. The proportion of critics making some use of the term 'titre' or its equivalents rose steadily through the nineteenth century, and by the 1860s and 1870s it had become the norm. Nearly 90% of critics reviewing the Salon during those two decades were writing about paintings at the Salon as having titles, with those titles given by the entries in the catalogue or a close variant. This was a general change and was not confined to critics of any particular aesthetic or political position.

³³ See for instance Marijke Jonker, 'Gustave Planche or The Romantic Side of Classicism', *Nineteenth-Century Art Worldwide*, 3/1, (2004). Jonker argues Planche has been seen by other scholars as more conservative than he actually was.

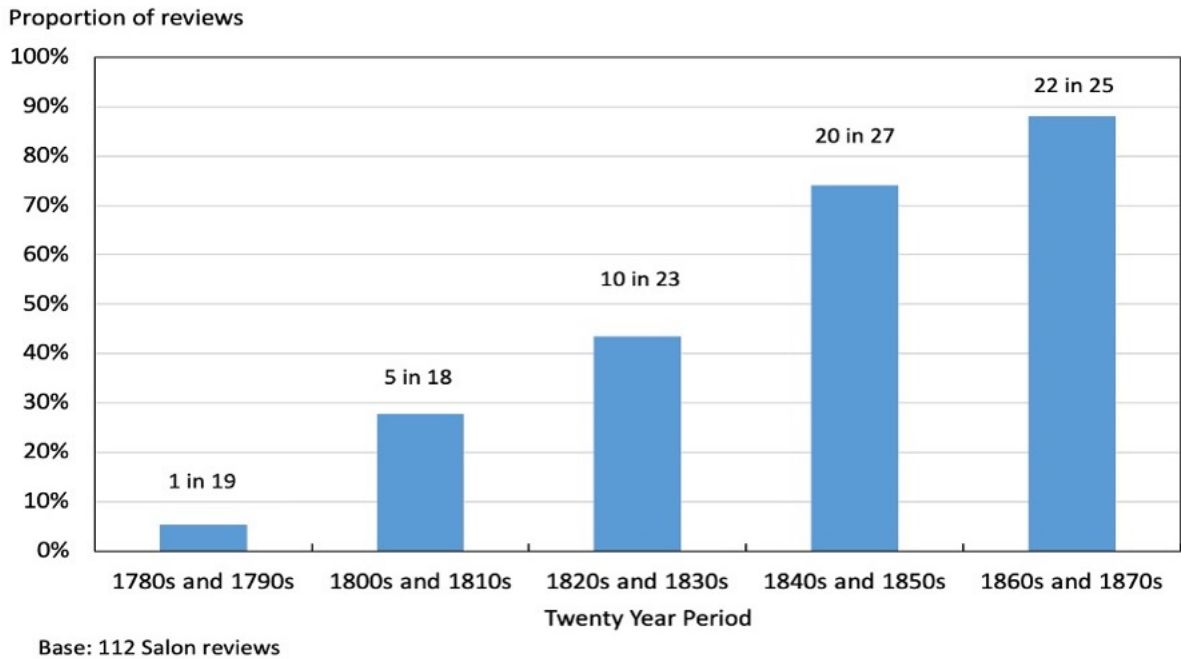


Chart 3.1: Proportion of critics using the term ‘titre’ or its equivalents in their Salon reviews, 1780s and 1790s to 1860s and 1870s.

The textual context of the language used by critics to refer to catalogue entries shows how the use of the term ‘titre’ developed and solidified to become a standard part of critical language. It also gives an idea of why critics may have started to think about catalogue entries that way. With the examples of art criticism in my sample for the decades from the 1780s to the 1820s, the convention typically followed when the catalogue entry was used in the main text was to relate it to the subject of the painting, as in phrases such as ‘un tableau de ...’ [a painting of ...], ‘un tableau representant ...’ [a painting representing ...], or ‘... est le sujet de’ [... is the subject of]. Critics who regarded the catalogue entry as giving the subject of the work were limited in how they could use it. They could do little more than adopt the standard critical device of disapprobation common in the eighteenth and early-nineteenth centuries of saying of a work that it was ‘not an X’, where X was the subject indicated by the catalogue entry.

In contrast, early occurrences of the use of ‘titre’ were almost all in contexts where the critic used the device of referring to a catalogue entry as a title and its relationship to the work not being that of a neutral or transparent expression of the subject matter to say something more about the work. In a review published anonymously of a seascape by Jean-Louis Demarne at the Salon of 1798 whose entry in the catalogue was ‘Vue des côtes de Normandie’ [View of the Normandy coast], François Joseph Aloyse Doix referred to the painting as ‘Celui qu’il donne sous ce titre simple et modeste’ [the one which he presents under this simple and

modest title] before going on to praise the work as being neither simple nor modest.³⁴ One of the stars and scandals of the Salon of 1819 was Théodore Géricault's *The Raft of the Medusa*. The political sensitivities of a work depicting survivors from the sinking of a French government frigate in the Atlantic Ocean in 1816 meant it was given the nonspecific generic entry of 'Scène de naufrage' [shipwreck scene] in the catalogue. The critic Auguste Jal was outraged by this act of censorship and scathingly referred to it in his review of the painting:

Permettez que je vous en fasse une description explicative; et sans m'arrêter à la faible raison qui a empêché l'artiste ou le directeur des musées de l'annoncer sous son véritable titre, celui de *Scène de naufrage du radeau de la Méduse*.³⁵

[Allow me to give you an explanatory description, and without dwelling on the feeble reason which prevented the artist or the Director of Museums from announcing it under its real title, that of *Shipwreck Scene of the Raft of the Medusa*.]

As the nineteenth century developed, critics of all persuasions were not only using the term 'titre' in contexts where they wished to draw attention to a catalogue entry to say something about the work or to suggest a more appropriate re-titling, but also as a standard way to describe catalogue entries. The increasing prevalence of referring to catalogue entries as titles would have acted as a signal of linguistic practice and would have reinforced this trend and solidified and conventionalised that use of language. Indeed, by the 1860s it had become the primary use of the term 'titre' in critical language. In his reviews of Salons from the mid-1860s published under the pseudonym Willem Bürger, Théophile Thoré made numerous references to catalogue entries being 'le titre du tableau' [the title of the painting] or 'son tableau intitulé ...' [his painting entitled ...].³⁶ The anonymous critic for the weekly periodical *L'Europe Artiste* referred to works being shown 'sous le titre ...' [under the title ...].³⁷ In the next decade, similar expressions were used frequently by the critic and member of the Academy Jules Claretie in his review of the Salon of 1876.³⁸ By that time, the understanding of catalogue entries as titles and the use of the term 'titre' and its equivalents to describe them had become an established and conventionalised part of critical discourse.

³⁴ Anon, *Itinéraire Critique du Salon de l'an VI*, (Paris: n. p., 1798), p. 18.

³⁵ Auguste Jal, *L'Ombre de Diderot et Le Bossu du Marais: Dialogue Critique sur le Salon de 1819*, (Paris: n. p., 1819), p. 124.

³⁶ Théophile Thoré, *Salons de W. Bürger 1861 à 1868*, vol. 2: 1864 - 1868, (Paris: Renouard, 1870).

³⁷ Anon, 'Salon de 1868', in *L'Europe Artiste*, 14 June 1868, 21 June 1868, 28 June 1868.

³⁸ Jules Claretie, *L'Art et les Artistes Français Contemporains avec un Avant-Propos sur le Salon de 1876*, (Paris: Charpentier et Cie, 1876).

The form and content of catalogue entries

Changes in the form and content of catalogue entries can also be interpreted in terms of the emergence of titling. However, as artists did not refer to or use their catalogue entries elsewhere in the catalogue itself this is not as straightforward as it is with reviewing critical language. It is also not a matter of looking at each entry and attempting to decide whether it is a title or not. The same piece of text could feature in a catalogue entry from the late eighteenth century, when it would not have been understood as a title, and one from the late nineteenth, by which time it would most likely have been considered as such. Rather, I approached this question by looking for the sort of changes in catalogue entries that might be associated with thinking about and using them as titles rather than as descriptions or classifications of the subject matter or the material object.

As I discuss in the Introduction, scholars have identified three main functions of the title.³⁹ The first and basic function is nomination - titles are names. To give a name to something is to individuate and identify it, to single it out from other things of the same type. This suggests the adoption of titling would be associated with a trend towards the individuation of catalogue entries and the objects they referred to, and, indeed, such patterns can be observed in the form of the entries in the catalogue. In Chart 3.2 I present, for each decade from the 1790s to the 1870s, the proportion of catalogue entries in the catalogues of the Salon relating to more than one object. What this shows is that from a situation in the 1790s where over one in five catalogue entries related to more than one object, there was a rapid decline in that form of use and by the 1870s effectively every painting on display at the Salon had its own entry.

³⁹ Gerard Genette, (1988), and Josep Bepa Camprubi, (2002).

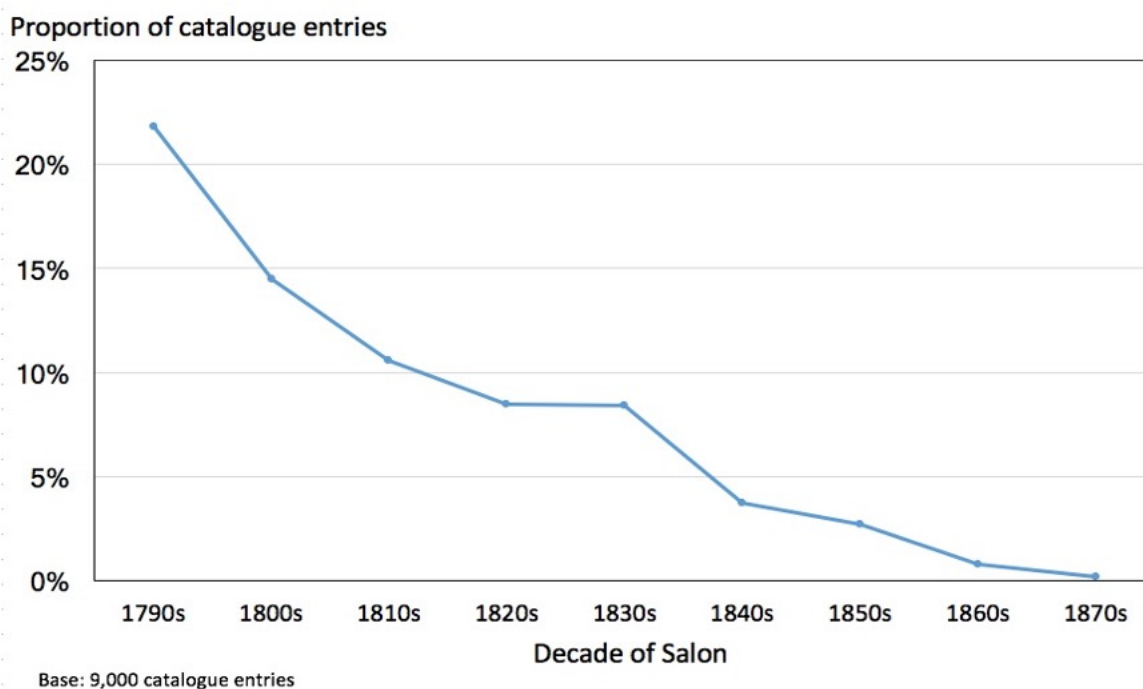


Chart 3.2. Proportion of entries used to refer to more than one object, 1790s to 1870s.

Additional evidence of the move towards individuation with catalogue entries comes from an examination of the language used in them. Artists sometimes used the terms ‘autre’ [another], as in ‘Autre paysage faisant pendant’ [Another landscape comprising a pendant] from the Salon of 1808, ‘idem’ [The same as the preceding] or in shortened form ‘id’, as with ‘Idem d’enfant’ [The same of a child] from the Salon of 1843, in entries to indicate the same type of work or subject as the previous entry. Uses of these terms in such contexts are also signs of non-individuation. Over the decades from the 1790s to the 1840s these kinds of term appeared in between 6% and 10% of main entries. By the 1860s their use had died out completely.

The process of individuation of catalogue entries represented a wholesale shift in the way in which art works were conceived. By the 1870s effectively all the paintings on display at the Salon were presented as individual objects in their own right and with their own names. To conceive of paintings in such terms can be seen as integral to the increasing commercialisation of the nineteenth-century French art world. As the century progressed the Salon became increasingly competitive as the number of artists rose more rapidly than the opportunities available to them to show paintings at the event. Market mechanisms including auctions and dealer sales also came to play a greater role in the sale of new art, and paintings changed hands more often.⁴⁰ Paintings became commodities, with each work of art an

⁴⁰ Developments in the production, distribution and exchange of new paintings in the nineteenth-century French art world are discussed in: Steven R. Adams, “Noising things abroad”: Art, Commodity and

individual object of economic value, and, as we have seen, individuated through having its own name. The assignment of an entry to effectively every painting on display at the Salon also supported the commodification of all types of painting. As Nicholas Green has discussed, in the second half of the nineteenth century items that were previously regarded as secondary such as the sketch were increasingly sold and collected in their own right.⁴¹

The second function of the title is to act as a commentary, saying something about the work it names, and so functioning semantically to contribute to the meanings it is given. The third function is to be 'seductive', where the title is used to stimulate the interest of the reader. Descriptive or classificatory entries can act as commentaries and may seduce the reader. However, as I will now come on to discuss, other uses of catalogue entries that can be associated more with thinking of them as descriptive or classificatory and less as titles had largely disappeared from the catalogue by the 1870s.

One indicator of the decline in thinking of catalogue entries as descriptive or classificatory is that as the nineteenth century progressed main entries largely ceased providing a reference to the medium or type of work, or to the physical object. The terms describing the medium or type of painting, for instance as a watercolour, drawing, pastel, enamel, or miniature, changed position in the catalogue to feature in a separate unit of text, following on from the main entry. Main entries also no longer referred to an object representing something, or to a frame containing something. To measure this trend, I looked at the proportion of main entries including a reference to the medium or type of work or the words 'representant' [representing], 'cadre' [frame], 'renfermant' [containing] or 'contenant' [containing]. The results are presented in Chart 3.3 which shows that such uses fell rapidly in the early nineteenth century and then declined further in the middle decades to feature in only 1% of main entries in the 1870s.

Commerce in Post-Revolutionary Paris', *Nineteenth-Century Art Worldwide*, 12, (2013), Jan Dirk Baetens, 'Vanguard Economics, Rearguard Art: Gustave Coûteaux and the Modernist Myth of the Dealer-Critic System', *Oxford Art Journal*, 33/1, (2010), pp. 27 - 41, Véronique Chagnon-Burke, 'Looking at and Buying Contemporary Art in Mid-Nineteenth-Century Paris', *Nineteenth-Century Art Worldwide*, 11, (2012), Nicholas Green, 'Dealing in Temperaments: Economic Transformation of the Artistic Field in France during the Second Half of the Nineteenth Century', *Art History*, 10/1, (1987), pp. 59 - 78, Nicholas Green, 'Circuits of Production, Circuits of Consumption: The Case of Mid-Nineteenth-Century French Art Dealing', *Art Journal*, 48/1, (1989), pp. 29 - 34, Nicholas Green, *The spectacle of NATURE. Landscape and bourgeois culture in nineteenth-century France*, (Manchester: Manchester University Press, 1990), pp. 17 - 40, and Linda Whiteley, 'Art et commerce d'art en France avant l'époque impressioniste', *Romantisme*, 40, (1983), pp. 65 - 76. Accounts of the economics of the art world include White and White, (1965), David Galenson and Robert Jensen, 'Careers and Canvases: The Rise of the Market for Modern Art in the Nineteenth Century', in *Current Issues in Nineteenth Century Art, Van Gogh Studies*, 1, (Waanders: Zwolle, 2007), and Monnier, (2004).

⁴¹ Green, (1987), p. 72.

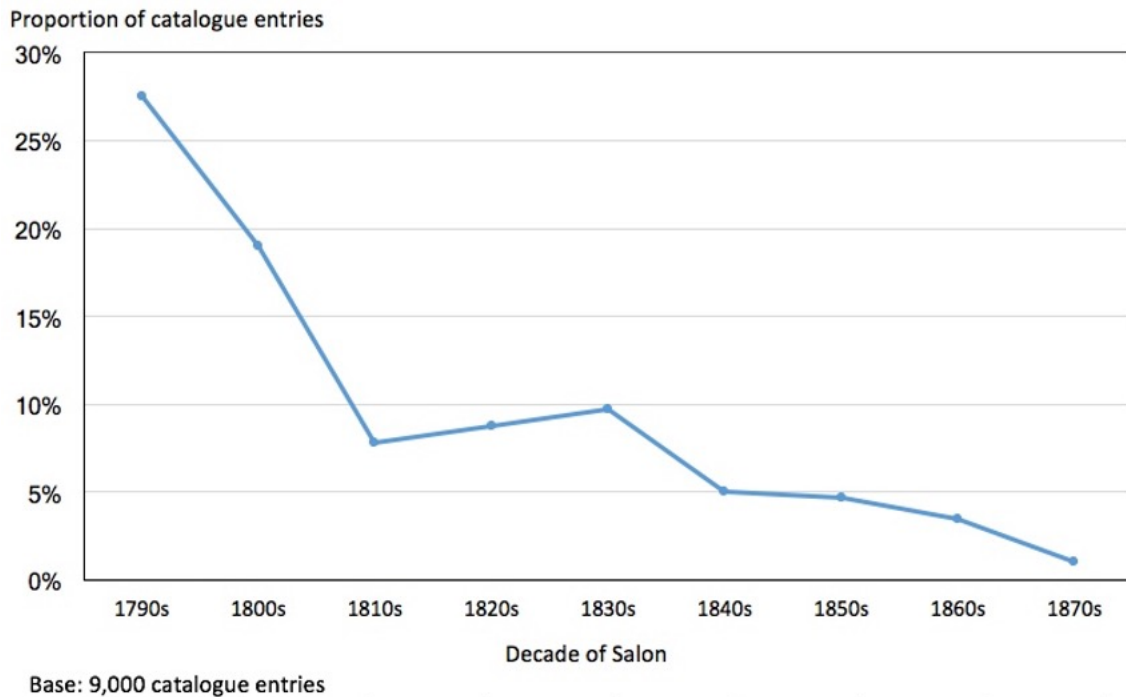


Chart 3.3. Proportion of main entries including a reference to the physical object or to the medium or type of work, 1790s to 1870s.

As I have discussed, Ruth Yeazell has noted that the basic function of the indefinite article is to classify. And in the early nineteenth century there was a sharp decline in their use at the start of catalogue entries such as 'Un portrait d'enfant' [a portrait of a child], as can be seen from Chart 3.4. From being used at the start of one entry in four in the 1800s, from the 1830s onwards only around one entry in fourteen began with the indefinite article. In parallel, with most generic categories there was a move away from the use of short classificatory entries such as 'portrait', or 'paysage', which had largely died out by mid-century.

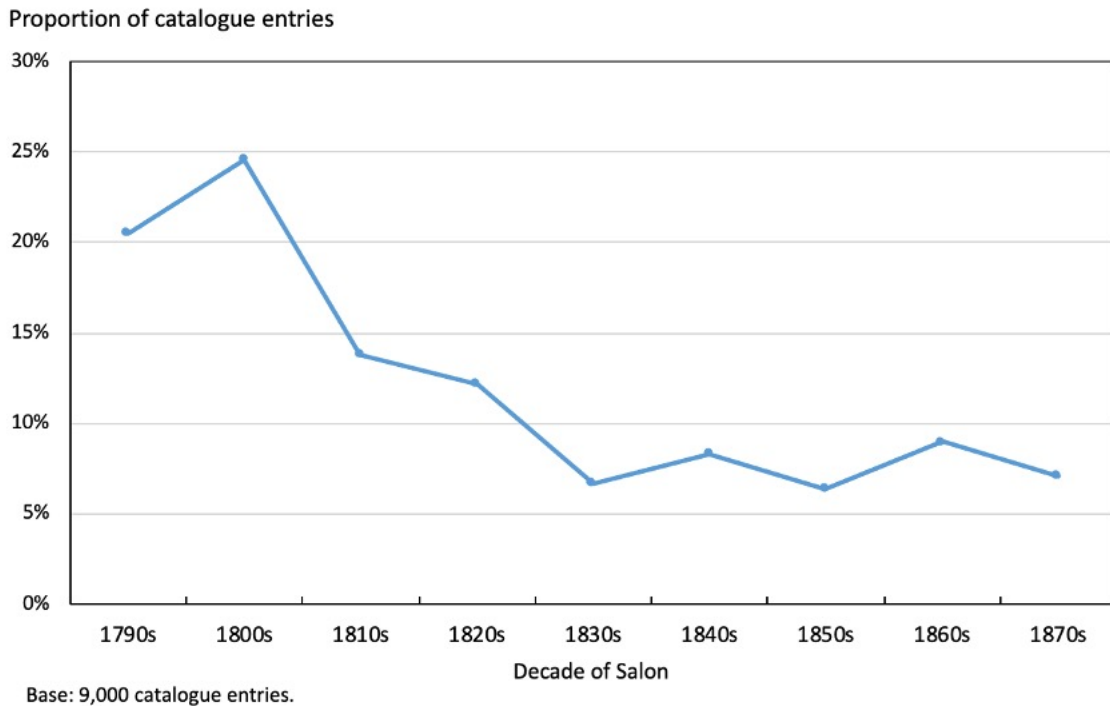


Chart 3.4. Proportion of main entries beginning with the indefinite article, 1790s to 1870s.

Artists also increasingly abandoned the extended descriptive entries often used in the 1790s and early nineteenth century to provide detail on the subject matter of paintings. This is shown in Chart 3.5, which looks at the proportion of all main entries for single oil paintings of 10 words or more in length. As can be seen the proportion of such entries of that length declined steadily from over one in four in the 1790s to only one in twenty in the 1870s. It also shows that in the 1870s fewer than 1% of main entries were of length 15 to 19 words, and very long main entries of 20 words or more had virtually disappeared from the catalogue. This compares with 8% and 5% respectively of main entries being of those word lengths in the 1790s.

The declining trend in the use of long descriptive entries was seen across all generic categories. With landscape entries artists became much less likely to use them to provide lengthy topographical descriptions of the location depicted, which often also included a specification of the place from which the view was taken. Fewer genre and historical entries set out in detail the actors and actions represented. In the middle of the nineteenth century the use of still-life entries to provide lengthy lists of all the items depicted also saw a sharp decline. With portrait entries, the practice of describing the background or the actions the sitter was engaged in effectively disappeared from the catalogue.

Proportion of catalogue entries

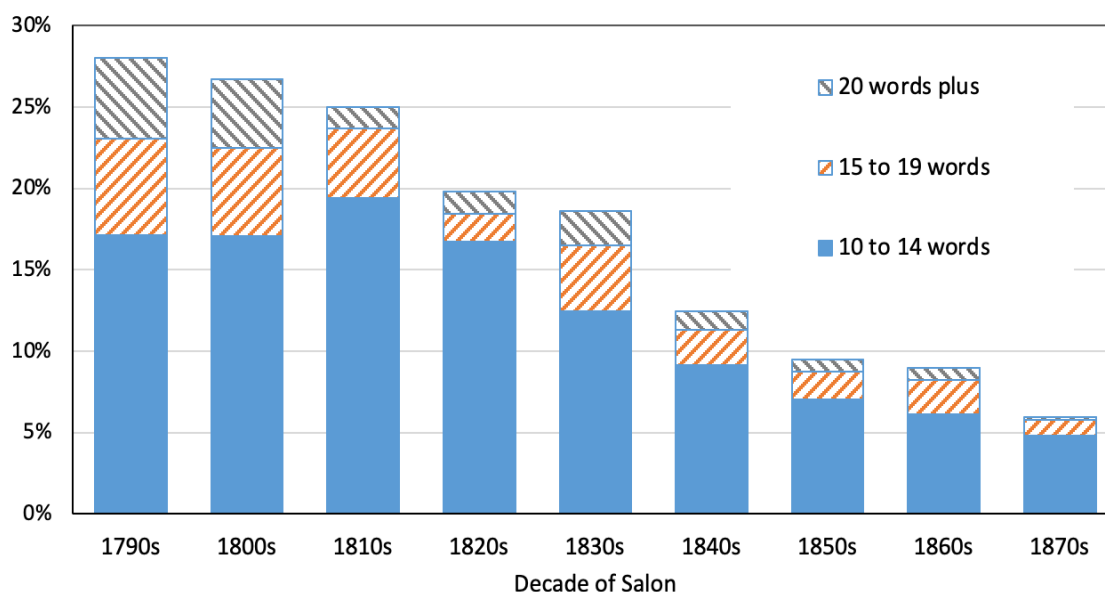


Chart 3.5. Proportion of all main entries for single oil paintings of length ten or more words, 1790s to 1870s.

We can also see the adoption of titling in the trend for artists to become much less likely to provide supplementary explication of the subject matter. By the 1870s the main entry itself had become the sole source of information from the artist to the reader of the catalogue on the subject of the painting. As can be seen from Chart 3.6, where I present the trend in this use of the catalogue, the proportion of entries including supplementary text fell steadily from just under one in five in the 1800s and 1810s to around 4% in the 1860s and 1870s.⁴² Some of this decline would have been associated with the move away from historical subjects in the nineteenth century, as entries for such works were the most likely to have supplementary text, but it was a trend seen across all generic categories.

⁴² The lower figure for the 1790s than the next two decades reflects the popularity during that decade of pendants, where the associated main catalogue entries often had detailed descriptions of the subject matter. Around 9% of catalogue entries in the 1790s related to pendants, compared with 2% in the 1800s.

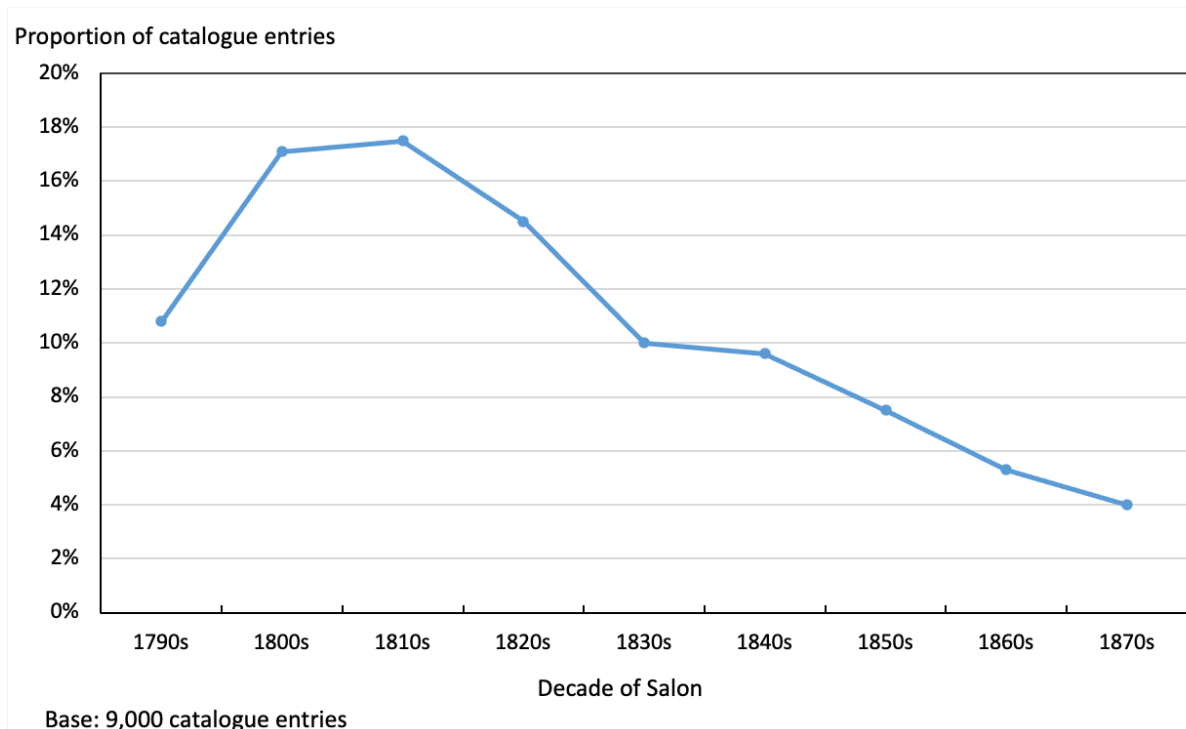


Chart 3.6. Proportion of entries with supporting explicative text, 1790s to 1870s.

What largely replaced the long descriptive and short classificatory entries were main entries of two types, each of which drew on functions of the title. Firstly, short entries that compressed meaning and commented on the work they named through succinctly directing the viewer's attention upon the most important elements of the painting including the individuals, objects, locations, or events depicted. Secondly, and at the same time, there was an increase in the use of entries that did not directly relate to the subject matter of the painting but were humorous, sentimental, suggestive, or simply puzzling in the way in which they alluded to the theme of the work. These allusive entries were also often short phrases. Allusive entries saw substantial growth in the nineteenth century and were predominantly associated with genre paintings. The critic Charles Baudelaire wrote about this trend in several of his *Salons*. Revealingly, he refers to the catalogue entries as 'titres', as he wished to distinguish what the entry is saying from the subject of the painting. In his Salon of 1859, for instance, he begins one section 'Mon cher M***, si j'avais le temps de vous égayer, j'y réussirais facilement en feuilletant le catalogue et en faisant un extrait de tous les titres ridicules ...' ['My dear M***, if I had the time to cheer you up, I could easily succeed in this by flipping through the catalogue and extracting all of the ridiculous titles ...].⁴³ It is hard to be precise, but my estimate is that

⁴³ Charles Baudelaire, *Ouvres Complet de Charles Baudelaire, II Curiosités Esthétiques*, (Paris: Michel Lévy Frères, 1868), p. 254.

over 50% of entries in the genre category during the 1850s to 1870s were of the kind Baudelaire so deplored. This compares with around 25% in the 1790s to 1810s.

One overall measure indicative of the adoption of succinct or allusive entries is the trend in the proportion of catalogue entries for single oil paintings of length three words or less, which I have presented in Chart 3.7. In the first few decades of the nineteenth century growth in short title-like entries was counter-balanced by a decline in short classificatory entries and in short entries supported by explicatory text. However, from the 1840s the balance shifted in favour of the former and the proportion of main entries of length three words or fewer began to increase, rising, as can be seen from Chart 3.7, from just under 20% in earlier decades to over one in three in the 1860s and 1870s. This was a trend seen across all generic categories bar portraits. For instance, the proportion of all historical entries of length three words or fewer rose from 22% in the 1790s to 38% in the 1870s. Over the same period, the corresponding proportion of all genre entries rose from 30% to 60%.

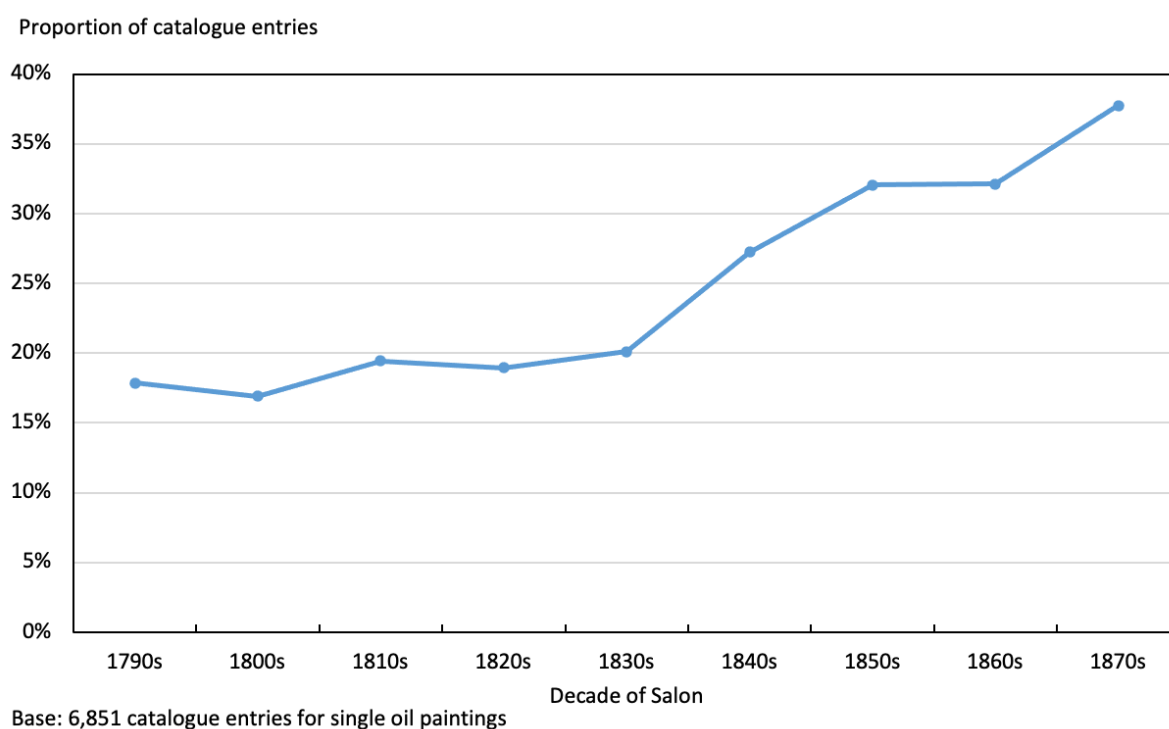


Chart 3.7. Proportion of main entries for single oil paintings of length three words or fewer, 1790s to 1870s.

Even if not always understood in the early nineteenth century as titles, succinct or allusive entries would have been of benefit to the artist in getting themselves noticed in the increasingly competitive environment of the Salon and the French art world. Artists were increasingly using

main entries to 'seduce' the reader of the catalogue. In much the same way as the use of the term 'titre' became standardised in critical language, the increasing prevalence of these types of entry in the catalogue would have reinforced those trends and solidified and conventionalised their use.

To illustrate the overall effect of the changes I have interpreted in terms of the emergence of titling, two extracts from Salon catalogues of the early- and late-nineteenth century can be compared. As with the extract from the catalogue of the Salon of 1806 given in Figure 3.2, in the late-eighteenth and early-nineteenth centuries the entries in the catalogue were quite heterogeneous, often not self-contained, and were in most cases descriptive or classificatory. By the 1870s entries had become much more homogeneous in form and in content, with, as can be seen from Figure 3.3 which presents an extract from the catalogue for the Salon of 1879, almost all items on display having their own self-contained and typically succinct entry. By that time main entries were being used in ways that I interpret as saying they were predominantly being thought of and used as titles.

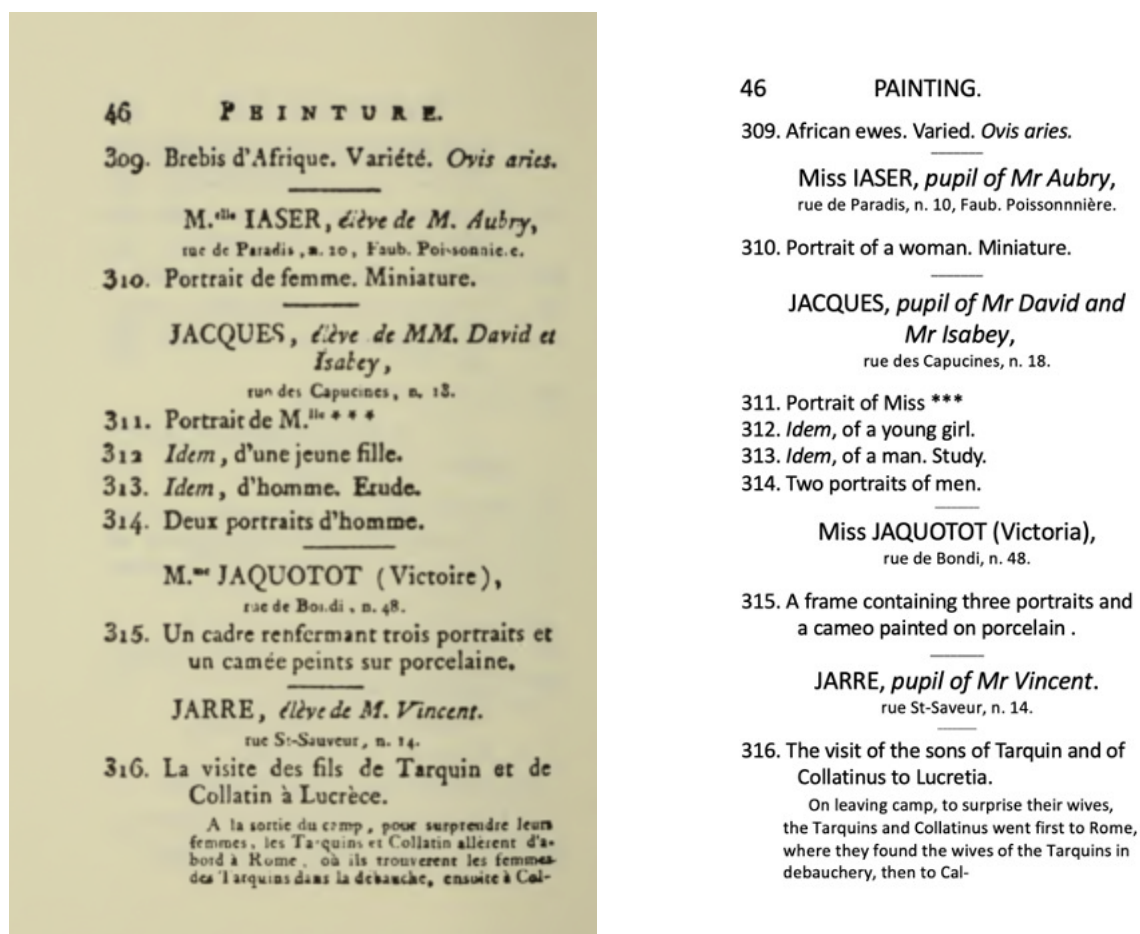
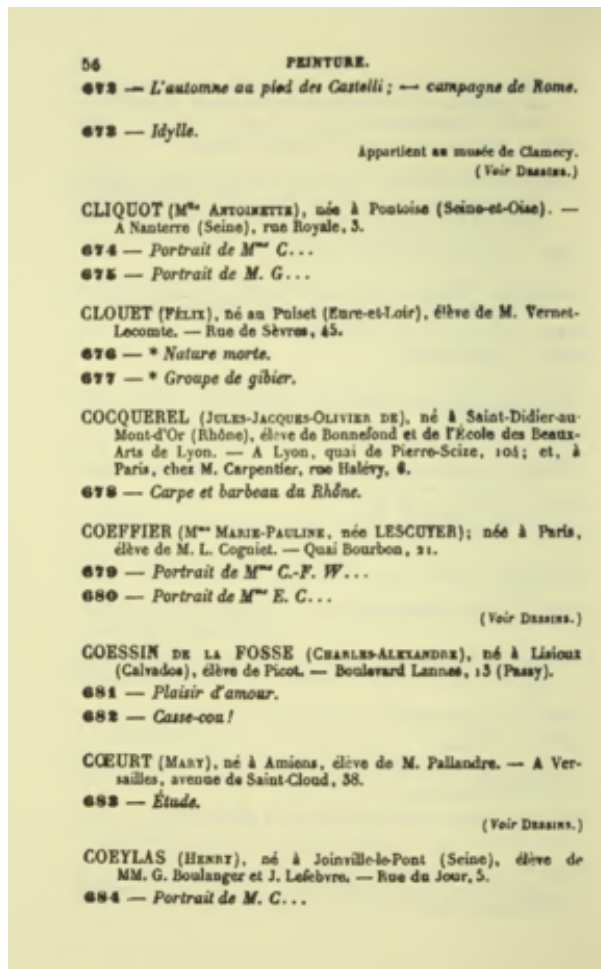


Figure 3.2. Extract from the catalogue for the Salon of 1806.



56

PAINTING.

672 - *Autumn at the foot of the Castelli; - Rome countryside.*
673 - *Idyll.*

*Belongs to the museum of Clamecy.
(see drawings)*

CLIQUOT (Miss ANTOINETTE), born at Pontoise (Seine-et-Oise). -
Nanterre (Seine), rue Royale, 3.

674 - *Portrait of Mrs G. ...*
675 - *Portrait of Mr G. ...*

CLOUET (FÉLIX), born at Poiset (Eure-et-Loir), pupil of Mr . Vernet-
Lecomte. - Rue de Sèvres, 45.

676 - * *Still-life.*
677 - * *Group of game.*

COCQUEREL (JULES-JACQUES-OLIVIER DE), born a Saint-Didier-au-
Mont D'Or (Rhône), pupil of Bonfond and of the Lyon School
of Fine Art. - Lyon, Quai de Pierre-Seize, 104, and
Paris, care of Mr Carpentier, rue Halévy, 6.

678 - *Carp and barbel of the Rhone.*

COEFFIER (Mrs MARIE-PAULINE, née LESCUTER); born in Paris,
pupil of Mr L. Cogniet. - Quai Bourbon, 21.

679 - *Portrait of Mrs C.-F. W. ...*
680 - *Portrait of Mrs E. C. ...*

COESSIN DE LA FOSSE (CHARLES-ALEXANDRE); born in Lisoux,
(Calvados), pupil of Picot. - Boulevard Lannes, 13 (Passy).

681 - *The Pleasure of Love.*
682 - *Daredevil!*

COEURT (MARY); born in Amiens, pupil of Mr Pallandre. -
Versailles, avenue de Saint-Cloud, 38.

683 - *Study.*

COEYLAS (HENRY); born in Joinville-le-Pont (Seine), pupil of
Mr G. Boulanger and Mr. J Lefebvre. - Rue du Jour, 5.

684 - *Portrait of Mr C.*

Figure 3.3. Extract from the catalogue for the Salon of 1879.

In summary, my work confirms Ruth Yeazell's tentative observation that the modern notion of the title of a painting started to get some traction in the French art world in the 1790s. It also challenges those art historical accounts of the nineteenth-century French art world in which catalogue entries are written about as if they were always used and understood as titles.

3.5 Catalogue Entries and Genre

The nineteenth-century French art world saw significant change in the prevalence of paintings in different genres at the Salon. Although the degree of liberality of the Salon jury and other factors such as limits on the number of paintings that could be submitted affected the proportions of entries in each genre at each Salon, the underlying trends are clear from my sample of catalogue entries. As can be seen from Table 3.3 the shares of entries in the historical category declined significantly from the 1820s onwards. Portraiture's share also declined substantially, in part, as Dominique de Font-Réaulx has examined, as a result of

technological innovation with the arrival of commercial photography at the end of the 1830s.⁴⁴ On the other hand, everyday genre paintings and landscapes appeared relatively more often at the Salon, with over 60% of entries for Salons from the 1850s to 1870s in those two categories, compared with 40% in the 1790s to the 1810s. Oil paintings with still-life subjects were also more common in the 1850s to 1870s than in earlier periods.

	Historical	Genre	Landscape	Portrait	Still-life	Other
1850s to 1870s	14%	26%	36%	17%	6%	1%
1820s to 1840s	21%	18%	34%	23%	3%	1%
1790s to 1810s	25%	15%	25%	30%	3%	1%

Table 3.3. Proportion of entries for single oil paintings in each generic category, 1790s to 1810s, to 1850s to 1870s.

Catalogue entries in each genre would have been used by artists to indicate to viewers of their works their sources of artistic value, how they were to be seen or appreciated, and to what they should be compared. The language used in catalogue entries would also have had meaning in relation to the institutional structures of the French art world and the broader socio-political environment. I have already mentioned how catalogue entries were instruments of consecration, and the sensitivities around the cataloguing of Géricault's *Raft of the Medusa*. I will now look at each of the generic categories in this way, giving readings of how the language used in catalogue entries functioned culturally and institutionally, and contrasting what that says with the work of other art historians who have written on the nineteenth-century French art world. For completeness, I have looked at all generic categories, although for some there is little that text mining reveals that is relevant to the questions I consider in this section.

Still-life Entries

A preliminary examination of my dataset suggested there were some strong trends in the use of different types of entry in the still-life category. However, as Table 3.3 indicates, still-life entries made up only a small proportion of all entries for single oil paintings in the Salon catalogue, and the number of entries in my core sample was not large enough for a robust analysis of their contents. To do this, I mined the full text of all the catalogues, not just those included in my sample, to identify entries in the still-life category. I searched for lines in those

⁴⁴ Dominique de Font-Réaulx, *Painting and Photography: 1839 - 1914*, (Paris: Flammarion, 2012), pp. 142 - 175.

text files that included the words 'morte' [dead] (for nature morte), and 'fleur' [flower] and/or 'fruit' [fruit]. I also looked for lines that included the most common types of flower or fruit including 'marguerite' [daisy], 'rose' [rose], 'raisin' [grape] and 'pêche' [peach]. Similarly, I identified lines containing the most commonly occurring other still-life subjects including 'gibier' [game], 'perdix' [partridge], 'légume' [vegetable], and 'hûître' [oyster]. Having identified a long list of potential still-life entries for each year, I removed all of those that were clearly not in the category. For instance, the word 'fleur' [flower] might feature in an entry for a painting of a flower market or a flower seller, or in the record of the prizes won by the artist at previous Salons where 'fleurs et fruit' [flowers and fruit] was one of the prize categories, and the eleventh-century saint 'Marguerite d'Écosse' [Margaret of Scotland] was a popular subject of historical paintings. I also extracted all entries which indicated that the medium was not oil. There are some still-life subjects not in my lists, the low quality of the scanning with some catalogues, and my only capturing the line of text in which the terms appeared means I will have missed some entries in the still-life category and may have misclassified others. However, comparing the list of entries compiled through text mining with two of the catalogues in pdf format indicated this approach identified around 90% of all still-life entries for paintings in oil along with the subjects indicated in those entries.

I then classified each still-life entry according to the combinations of subject matter indicators used in it. I wanted these to reflect artistic distinctions that held in the nineteenth century French art world and so I turned to Lévesque and Watelet's dictionary of the arts.⁴⁵ In the dictionary, they distinguish 'tableaux de nature morte' [paintings of nature morte], which are paintings of literally 'dead nature', such as game, fish, or vegetables, from 'tableaux de fleurs' [flower paintings] and 'tableaux de fruits' [fruit paintings].⁴⁶ This was both a taxonomic and an aesthetic distinction. With a painting of 'nature morte' the aesthetic interest arose solely from its verisimilitude, whereas paintings of flowers and fruit were in addition to be pleasing to the eye. I also wanted the categories to be those whose levels of use saw significant change during the nineteenth century and so might raise questions of art historical interest.

These considerations led me to classify still-life entries into four distinct categories. The first category was entries whose only indications of subject matter were 'fleurs' [flowers] and/or 'fruits' [fruits], or included specific types of flower or fruit, either on their own or in combination with 'fleurs' or 'fruits', for instance 'fleurs et raisins' [flowers and grapes]. I have labelled this category 'Fleurs or Fruits only' in Chart 3.8. All other still-life entries included a reference to

⁴⁵ Lévesque and Watelet, (1792), Vols 1 to 4.

⁴⁶ Lévesque and Watelet, (1792), Vol. 3, p. 180.

subjects other than flowers of fruit. They have been grouped together in Chart 3.8, and in Chart 3.9 I have split them into three types. The first was entries including the term 'nature morte', either on its own or with indicators of specific 'dead nature' subjects such as 'nature morte, gibier' [nature morte, game]. I have labelled these entries 'Nature Morte' in Chart 3.9. The second category, labelled 'Dead Nature' in Chart 3.9, was entries that only included indicators of specific 'dead nature' subjects such as 'gibier et légumes' [game and vegetables]. The final category was for entries indicating subjects including a combination of flowers or fruit, and dead nature, either generically through 'nature morte', as with 'fruits et nature morte' [fruit and nature morte], or specifically, for instance 'fleurs et perdix' [flowers and partridge] or 'gibier et pêches' [game and peaches]. This category is labelled 'Fleurs/Fruits and Nature Morte or Dead Nature' in Chart 3.9.

The trends in the relative levels of use of each kind of still-life entry are presented in Charts 3.8 and 3.9. Chart 3.8 compares 'Fleurs or Fruit only' still life entries to all other kinds of entry, and Chart 3.9 looks in more detail at those other kinds of entry, presenting their relative levels of use. I have deviated from the usual format with the charts presented in this chapter of looking by decade, as the levels of use changed substantially from one year to the next in the 1870s. The mix of still-life entries in the catalogue for the Salon of 1870 were very similar to that with Salons in the 1860s, and I have combined 1870 and the 1860s together. There was no Salon in 1871, the year following the French defeat in the 1870 war with Prussia. The uses at the Salons in 1872, 1873 and 1874 were very similar, as were those at the Salons of 1875 and 1877, and the Salons of 1878 and 1879. Each of these three periods covering the years from 1872 to 1879 is presented separately in Charts 3.8 and 3.9.

The absolute number of still-life entries of all kinds identified in my text mining can be compared with the total number of entries in the catalogue for oil paintings to give an indication of the prevalence of still-life painting in oil at the Salon.⁴⁷

⁴⁷ I have estimated the total number of entries in each Salon though multiplying the number of entries under painting in the catalogue with the proportion of entries for paintings in oil indicated by my dataset for that Salon or for all of the Salons in the same decade if it was not represented in my dataset.

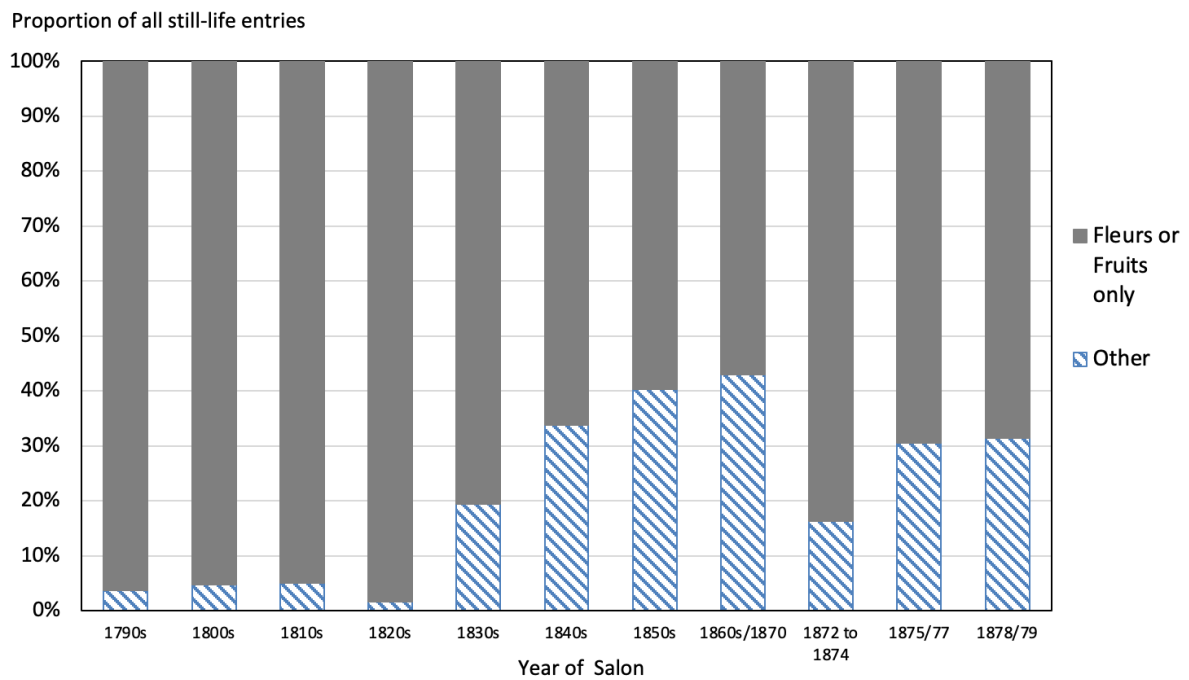


Chart 3.8. Still-life entries whose only indications of subject matter were of flowers or fruit and other still-life entries , proportion of all still-life entries for single oil paintings, 1790s to 1870s.

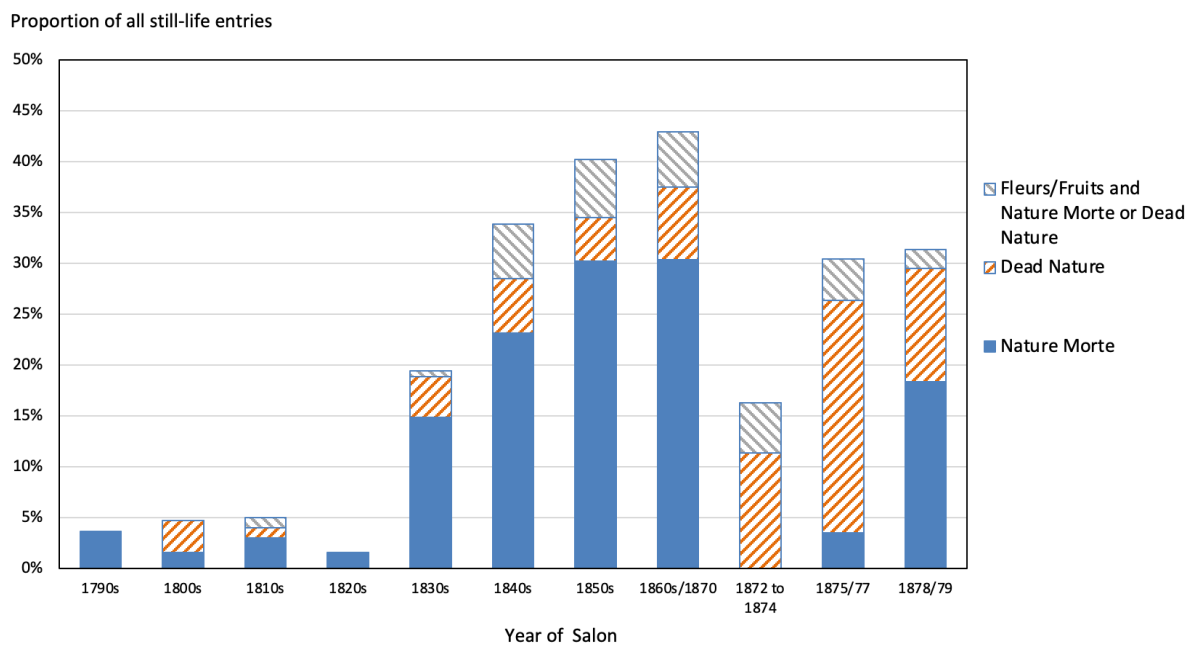


Chart 3.9. Still-life entries including a reference to subjects other than flowers or fruit, proportion of all still-life entries for single oil paintings, 1790s to 1870s.

I will now give my reading of these trends. As can be seen from Chart 3.8, in the decades from the 1790s to the 1820s the display of still-life subjects at the Salon was dominated by paintings of fruits and flowers. At the Salons during those decades 95% or more of still-life entries for oil paintings in oil indicated those subjects. In absolute terms, there were only 16 oil paintings of dead nature displayed in total at the 18 Salons in those four decades, compared to over 400 oil paintings of flowers and fruit.

Looking at the absolute numbers of still-life entries, confirms the overall trend given in Table 3.3 that from the 1840s onwards still-life subjects became increasingly more common, rising to account for around 6% of all single oil paintings in the 1860s compared with 3% in the 1830s. The art historian John McCoubrey has reviewed the painting of still-life subjects in the nineteenth-century French art world.⁴⁸ He ascribes the rise in popularity of still-life subjects in the middle decades of the century to a revival of interest and critical re-appraisal and promotion of the work of the eighteenth-century artist Jean-Baptiste Simeon Chardin. A renewed interest in Dutch art from the seventeenth century, with its strong tradition of still-life painting, may also have been an influence. The art historian James Kearns positions the growth as part of cultural and philosophical trends towards positivism and naturalism, with still-life seen as a form of scientific enquiry and an expression of artistic sincerity, and as being promoted by aesthetic theories privileging formal values over subject matter.⁴⁹

Neither McCoubrey nor Kearns look at the mix of still-life subject displayed at the Salon. What Chart 3.8 suggests is that from the 1830s to the 1860s a growing proportion of still-life paintings in oil included subjects other than flowers or fruit. Within that overall trend, as shown in Chart 3.9, there was more use of entries that did not include the term 'nature morte' and those with mixed indicators of dead nature and fruits or flowers. Entries in that latter category such as 'gibier, fleurs et fruits' [game, flowers and fruit] from the Salon of 1868 suggest the distinction in Academic doctrine between paintings of 'nature morte' and of flowers and fruit as set out in their dictionary by Lévesque and Watelet was beginning to be eroded. Some artists were combining both types of subject matter in their paintings of still-life subjects and were signalling that through their catalogue entries. Although the number of entries is small, some artists were using the term 'nature morte' as a prefix followed by indicators of both types of subject matter, as with 'nature morte: fleurs, fruits et animaux' [nature morte: flowers, fruits,

⁴⁸ John W. McCoubrey, 'The Revival of Chardin in French Still-Life Painting, 1850-1870', *The Art Bulletin*, 46/1, (1964), pp. 39 - 44.

⁴⁹ James Kearns, 'No Object too Humble? Still-life Painting in French Art Criticism during the Second Empire', in Brian Rigby, (ed.), *French Literature, Thought and Culture in the Nineteenth Century*, (London: Macmillan, 1993), pp. 148 - 168.

and animals] from the Salon of 1847. This use of catalogue entries also raises the question of whether the Academic cultural category of 'nature morte' was itself being challenged and redefined to extend to all kinds of still-life subject.

Some support for my tentative reading of how catalogue entries functioned in relation to still-life painting comes from looking at the context of use of 'nature morte' in my sample of critical writing for the period from the 1830s to the 1860s. Although most of the reviews that looked at still-life painting had separate sections devoted to reviews of 'nature morte' and of paintings of flowers and fruit, some critics were using it in the broad sense to include all types of still-life subject. In the *Le Salon de 1833* by Gabriel Laviron and Bruno Galbacio, the section entitled 'peinture de nature morte' [nature morte painting] includes reviews of paintings of fruit and flowers as well as those of game, lobsters and oysters.⁵⁰ In his *Salon de 1847* the critic Théophile Thoré characterises the artist Philippe Rousseau as a painter of 'nature morte' before discussing paintings of game and flowers by the artist.⁵¹ In his reviews of the Salons from the 1850s, the critic Jules-Antoine Castagnary also used 'nature morte' to refer to all types of still-life subject matter.⁵²

Some critics reflected in their reviews on these changes in the scope of use of 'nature morte' for the painting of still-life subjects. As well as promoting a unified category of still-life painting, in his review of the Salon of 1864, Thoré, was concerned that 'nature morte' was not an appropriate name as it did not adequately describe the full range of still-life subject matter. He comments that 'passons à la «nature morte» à la *vie coye* (still-life), comme on disait bien mieux au seizième siècle et au dix-septième' [let us move on from "nature morte" to *vie coye* (still-life), as was much better said in the sixteenth and seventeenth centuries].⁵³ Elsewhere in his *Salons* the alternatives he promoted included the English term 'still-life' or 'nature immobile' [immobile nature]. Critics also had concerns or resisted these changes. In his review of Philippe Rousseau's painting *Fleurs d'Automne* at the Salon of 1866, the critic and administrator Charles Blanc adopts a dismissive tone in referring to 'La peinture de ce qu'on appelle avec plus ou moins de justesse «la nature morte» ...' [the style of painting called more or less aptly 'nature morte'].⁵⁴

⁵⁰ Gabriel Laviron and Bruno Galbacio, *Le Salon de 1833*, (Paris: Abel Ledoux, 1833), pp. 367 - 374.

⁵¹ Wilhelm Bürger, *Salons de T. Thoré, 1845, 1846, 1847, 1848*, (Paris: Jules Renouard: 1870), p. 501. Wilhelm Bürger was a pseudonym used by Thoré.

⁵² Jules-Antoine Castagnary, *Salons (1857 - 1870)*, (Paris: Charpentier, 1892).

⁵³ Théophile Thoré, *Salons de W. Bürger 1861 à 1868, 2 Vols*, (Paris: Jules Renouard: 1870), Vol. 2. p. 119.

⁵⁴ Charles Blanc, 'Le Salon de 1866', in *Gazette des Beaux-Arts*, XX, (1866), p. 534.

My account of the changing status of still-life subjects and the contested meaning of the term 'nature morte' problematises the readings given by McCoubrey and Kearns. Both treat still-life as having been a well-defined genre in the nineteenth-century French art world, signified in French by the term 'nature morte', and encompassing all types of subject in the still-life category. Other scholars have made the same assumption.⁵⁵ This is a topic meriting further investigation, and it would be of interest to look in detail and in quantitative terms at critical writing to see how terms such as 'nature morte' were being used. However, that is beyond the scope of this thesis as the small sample of critical writing I was able to compile in electronic format is not adequate.

If we move on to the 1870s, we can see from Chart 3.9 that in the Salons of 1872, 1873 and 1874 there was no use at all in the catalogue of the term 'nature morte'. Depictions of flowers and fruit also accounted for the large majority of still-life paintings in oil at the Salon, accounting for 85% of paintings on display. Such major changes in the language used in catalogue entries and in the proportions of different still-life subjects displayed at the Salon would have been the result of editorial pressure or censorship of artists' submissions, and of Salon jury policy in selecting paintings, rather than artistic choice.

The art historians Albert Boime, Patricia Mainardi and Jane Mayo Roos have all examined in detail the cultural politics and conflicts around the Salon during this period.⁵⁶ Following the formation of the Third Republic in 1870, the administration of the fine arts in France was reformed in 1871. Charles Blanc was given the key post of Director of Fine Arts. He was a staunch conservative in taste and for Blanc state sponsorship of the fine arts should support its didactic and elevating functions. With Blanc a powerful influence, and directly responsible for the appointment of 40% of its members, the juries for the Salons of 1872 and 1873 were draconian, in both years selecting only around 2,000 of the 5,000 works submitted. Its decisions met with a plethora of protests from artists, critics, and journalists. Blanc was replaced in 1873 by the Marquis de Chevennières, who served until 1877 and retained his predecessor's conservative cultural stance. The Directors of Fine Arts in 1878 and 1879 adopted a somewhat more liberal position regarding the rules and admission process to the Salon.

⁵⁵ See, for instance, Jo Briggs, 'Condemned to Sparkle: The Reception, Presentation and Production of Léon Bonvin's Floral Still-lives', *Oxford Art Journal*, 38/2, (2015), pp. 247 - 262.

⁵⁶ Albert Boime, 'The Salon des Refusés and the Evolution of Modern Art', *Art Quarterly*, 32, (Winter 1969), pp. 411 - 26, Mayo Roos, (1996), pp. 37 - 55, Mainardi, (1993).

All these authors look in detail at landscape painting, and the 1860s and 1870s as the critical period for the emergence of Impressionism. They note that under Blanc's directorship landscape painters were largely excluded from the jury, and the jury rejected many of the landscape paintings submitted for the exhibition. None look at still-life painting. However, it is likely that the use of the term 'nature morte' in the broad sense and the attempts by critics such as Thoré and Castagnary to construct a unified school of still-life painting under that name were seen by Blanc as challenges to the Academic aesthetic principles he espoused and to the established hierarchies he wanted to maintain. In appearing in the official catalogue of the Salon, entries including the term 'nature morte' might have signalled a degree of institutional recognition of these changes. Under Blanc, the term 'nature morte' disappeared from the catalogue and the display of still-life subjects reflected the more elevated status of depictions of flowers and fruit established in Academic doctrine. The more liberal policies of later Directors may be reflected in the increased presence at the Salon of still-lives with dead nature subjects, and the re-emergence of the term 'nature morte' in the catalogue. This is another topic that merits further investigation, and would be well-suited to conventional art historical methods such as detailed reviews of the archives of the Salon administration, critical journals and other relevant archival material.

Landscape Entries

As measured by the proportion of entries in the catalogue, landscape became the most common genre of oil painting at the Salon in the 1820s and remained so through to the 1870s. In looking at what the language used in catalogue entries tells us about landscape painting I will first consider the two principal sources of aesthetic value being signalled: the location represented, and the temporal or meteorological phenomena depicted. These elements were central to the conceptual and aesthetic framework for landscape painting given by Academic theory. The Academician Pierre-Henri de Valenciennes, in his 1800 advice to the landscape painter that remained influential throughout the nineteenth century, and Lévesque and Watelet, in their 1792 Dictionary, stress the need for landscape artists to be skilled in the representation of all the aspects of nature including its objects and its phenomena.⁵⁷ The

⁵⁷ Pierre-Henri de Valenciennes, *Éléments de Perspective Pratique, à l'usage des Artistes, suivis de Réflexions et Conseils à un Elève sur la Peinture, et particulièrement sur le genre de Paysage*, (Paris: Desenne, Duprat, 1800), pp. 404 - 407. Lévesque and Watelet, (1792), Vol. 4, pp. 34 - 35. For a detailed discussion of these theories of landscape and of the tensions and contradictions within and between them see Élisabeth Lavezzi, 'Typologie et Imaginaire du Paysage dans les Éléments de Valenciennes, Hérietier de Pies et Watelet', in Alain Grosrichard (ed.), *Dénouement des lumières et invention romantique: actes du colloque*, Genève, 24-25 November 2000, (Geneva: Librairie Droz, 2003), pp. 207 - 230.

principal choices to be made by the artist included the site to be depicted and the effects or accidents of nature they wished to capture. These phenomena were themselves highly conventionalised, being largely restricted to the four times of day (morning, mid-day, evening, and night) or the associated light such as sunrise, sunset or of the moon, to the four seasons, and to calm, stormy or foggy weather. Valenciennes devoted much of his advice to the aspiring landscape painter to the representation of these hours of the day and times of the year.

The trends in how landscape entries signalled one or both of these sources of aesthetic value can be read in terms of change in their importance and distinctiveness. The greater the proportion of landscape entries referring to a location the more important that source of value is within landscape painting. The same is the case for entries with a temporal or meteorological reference. The smaller the proportion of entries referring to both is relative to entries giving only one, the more those sources of aesthetic value are signalled as distinct. I have looked at these trends with the landscape entries in my dataset, which are of a sufficient number for this purpose. They are presented in Chart 3.10. As can be seen, throughout the period from the 1790s to the 1870s the large majority of landscape entries signalled one or both of these two sources of value.⁵⁸ In all decades from the 1800s onwards, over 75% of landscape entries included such references.

Chart 3.10 also shows that in the 1790s and 1800s, location on the one hand and the weather, hour of day or time of year on the other were often presented as distinct subjects in their own right, each capable of generating what critics would have called an overall 'effet' [effect].⁵⁹ At that time the proportions of entries with a meteorological or temporal reference only and a location only were much greater than those including both kinds of reference. Over the next three decades location grew to be by far the single most important source of aesthetic value within landscape painting as signalled through the entries in the catalogue. By the 1830s, nearly 90% of landscape entries for single oil paintings included a location. The growing importance of location displaced that of the weather, hour of day or time of year. In the 1820s and 1830s only 13% of landscape entries included such a reference, compared with 26% in

⁵⁸ The entries in the landscape category that did not include such a reference were primarily those giving a generic description as in 'paysage' [landscape] or 'marine' [seascape], or generic subjects without a specific location such as 'vaches à l'abreuvoir' [cows at the trough], 'lisière de bois' [edge of the wood], or 'maison de paysan' [peasant house].

⁵⁹ In Salon criticism the success of a landscape was often judged in terms of its 'effets' in the plural or its 'effet' in the singular. As Lévesque and Watelet discuss in detail in their 1792 Dictionary, each part of a painting was considered to have its own 'effet', or effect upon the viewer, which with landscapes could be that of an oak tree or the play of light upon the water. All the individual effects contributed to the overall effect of the work. Lévesque and Watelet, Vol. 2, (1792), pp.111 - 121.

the 1790s and 1800s. These natural phenomena were also much less likely to be presented as subjects in their own right. In the 1830s around 70% of entries including a meteorological or temporal reference also included a location, compared with 35% in the 1790s.

Proportion of catalogue entries

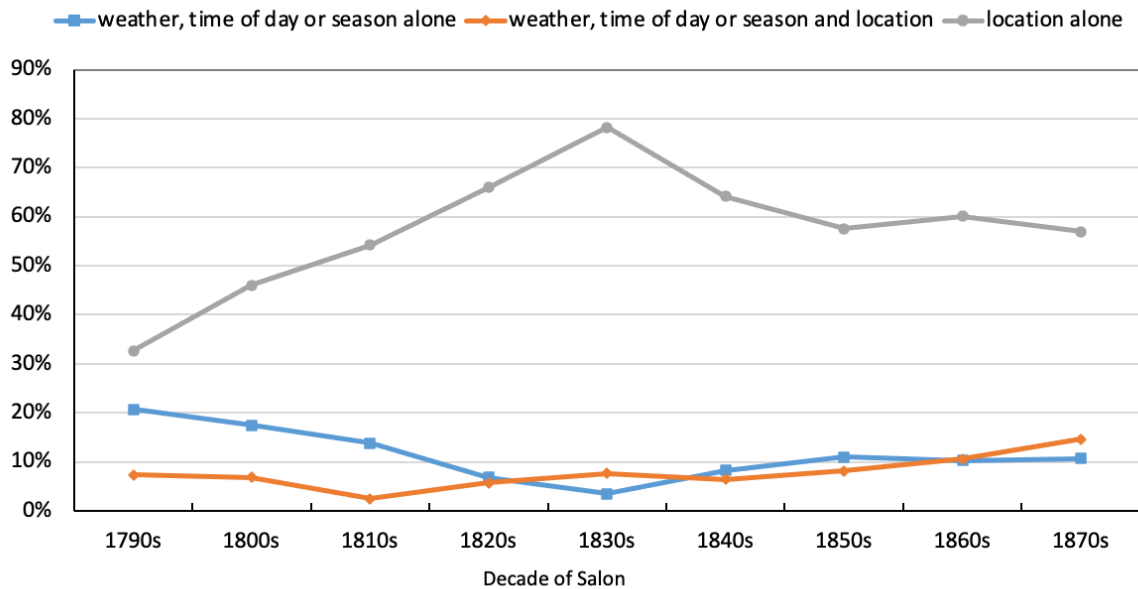


Chart 3.10. Proportion of landscape entries for single oil paintings containing a temporal or meteorological reference alone, a temporal or meteorological reference and a location, or a location alone, 1790s to 1870s.

Over the period from the 1840s to the 1870s there was a partial reversal of the trends seen in earlier decades. As Chart 3.10 shows, landscape entries signalled a growing importance of the temporal and meteorological, either as a subject in its own right or in combination with a location. Several scholars have explored how from the mid-century there was an increasing concern from landscape artists with accidental and transitory effects, with critics responding both positively and negatively to this trend.⁶⁰ Although there was a corresponding decline in location as a source of value, the catalogue entries indicate it remained important to landscape painting for artists to respond to somewhere in particular and to signal that response. Through to the 1870s the large majority of landscape entries included a reference to a location.

⁶⁰ For instance, see John House, 'Authority versus independence: the position of French landscape in the 1870s', in Richard Thompson, (ed.), *Framing France: Essays on the Representation of Landscape in France, 1870 - 1914*, (Manchester: Manchester University Press, 1998), p. 21.

If we now look at the locations given in landscape entries, we can see how text mining can complement and extend other art historical accounts of landscape painting in the nineteenth century French art world. As Nicholas Green has explored, during the middle decades of the nineteenth century the experience and conception of the French countryside saw major change.⁶¹ The domestic countryside was celebrated in a series of popular books. It became increasingly inter-twined with conceptions of national identity and national heritage. Nature tourism in which the countryside as such was to be experienced, as distinct from the trips to country houses or estates that were previously the norm, was starting to become part of middle-class Parisian life. The 1820s also saw the appearance of English landscape painters at the Salon, including John Constable and John Copley, and there was growing critical attention to Dutch painting. Landscape painters from both countries typically depicted their own countryside. At the Salon there was a corresponding increase in the depiction of locations in France. This is presented in Chart 3.11, which shows that domestic locations rose from appearing in 37% of landscape entries for single oil paintings giving a location in the 1790s to 76% in the 1830s.

As Albert Boime has explored, in the early to middle decades of the nineteenth century the reputation and status of classical Academic landscape painting was increasingly under threat, being seen by a growing number of critics as sterile and conventionalised.⁶² Rather than the ennobled representation of nature through the picturing of an idealised classical location, what mattered was an imaginative or poetic response to the natural world. Typically, these classical locations were associated with Italy, and as can be seen from Chart 3.11, from appearing in 43% of landscape entries in the 1790s, Italian locations declined to account for only 7% of locations in the 1830s.

⁶¹ Green, (1990), pp. 67 - 126.

⁶² Albert Boime, *Art in an Age of Counter-Revolution, 1815 - 1848*, (Chicago: Chicago University Press, 2004).

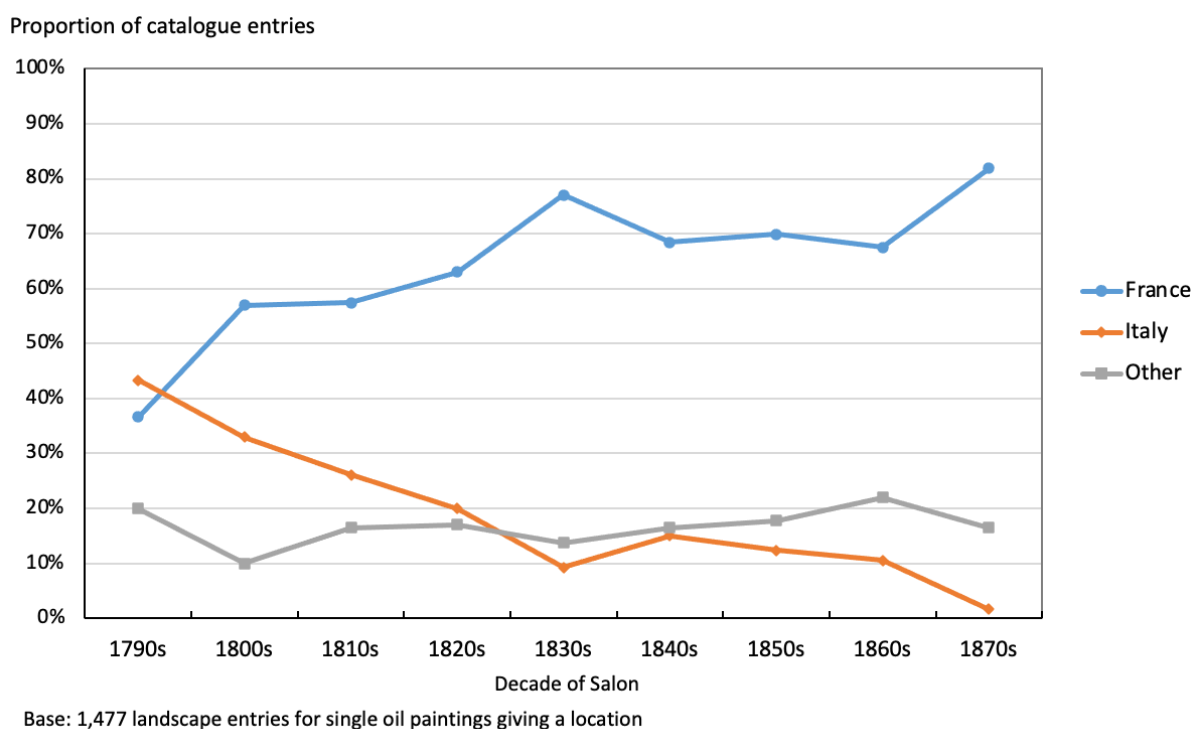


Chart 3.11. Proportion of landscape entries for single oil paintings containing a reference to a location, split by France, Italy and Other, 1790s to 1870s.

Chart 3.11 also shows that French locations remained dominant through to the 1870s. Diana Seave Greenwald has observed that the construction of a national rail system from the 1840s supported the growth in artist colonies in regions including Brittany and Normandy.⁶³ It also opened up the countryside to the rapidly growing urban middle-class, especially the environs of Paris and the northern coast, boosting demand for depictions of those locations. A few critics continued to maintain Academic principles and the historical or classicising landscape as the true measure of a landscape artist's worth. However, by the 1860s the idea of the classical landscape had been largely left behind by most critics and artists.⁶⁴ Amongst other critics, there were those who continued to argue for the primacy of the artist's expressive or poetic response to nature. Others advocated landscape painting that investigated nature objectively as it is or as it is seen or examined the 'modern' experience of life. However landscape painters saw themselves, or whatever movement or tendency they identified with, their catalogue

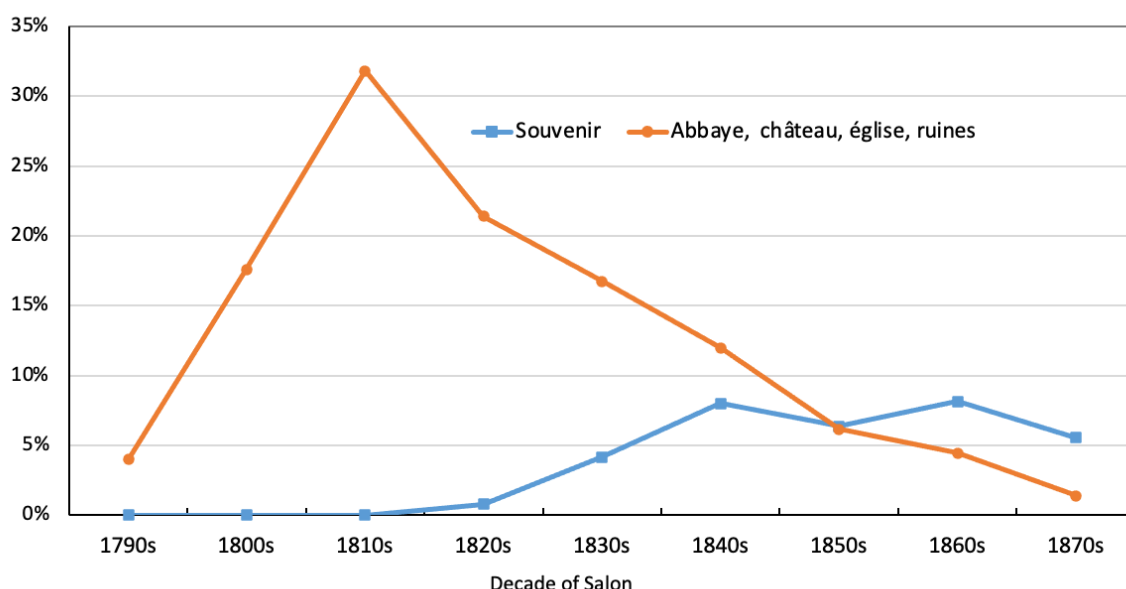
⁶³ Greenwald, (2019), p. 157.

⁶⁴ Sloane, (1973), pp. 100 - 108.

entries indicate their responses to the landscape were predominantly expressed through domestic locations.

The languages used in landscape entries can also be read as suggesting how conceptions of the domestic countryside changed during the nineteenth century. The term 'souvenir' [souvenir] can be read as signalling, at least in part, that the countryside was something to be experienced directly and valued in its own right. Use of the term first became significant in landscape entries giving a domestic location in the 1830s and rose to feature in around 6% to 8% of such entries in the decades from the 1840s to the 1870s. Chart 3.12 shows this trend. On the other hand, landscape entries suggest that the strength of the patrimonial conception of the domestic countryside first grew and then weakened substantially during the nineteenth century. These changes can be read as associated with the use of the terms 'abbaye' [abbey], 'châteaux' [castle or manor house], 'église' [church], and 'ruines' [ruins]. As Chart 3.12 shows, in the early decades of the nineteenth century such subject matter was an integral part of the increase in the depiction of French locations, rising from featuring in 4% of entries giving a domestic location in the 1790s, to over 30% of such entries in the 1810s. The use of these terms then fell away, to feature in only 2% of landscape entries giving a domestic location in the 1870s.

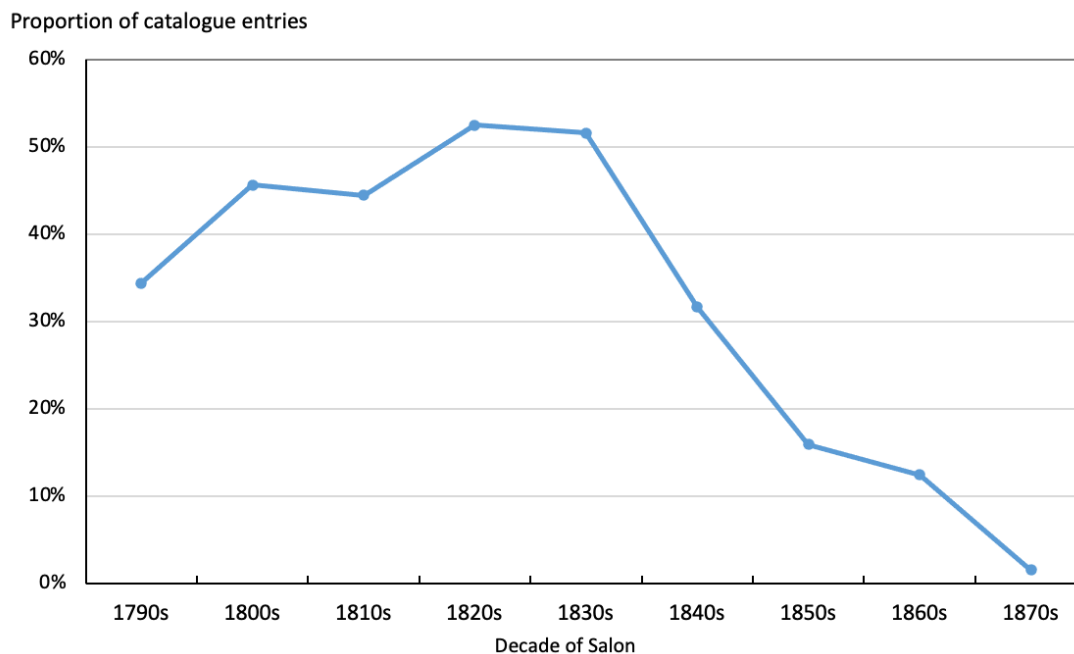
Proportion of catalogue entries



Base: 996 landscape entries for single oil paintings giving a French location

Chart 3.12. Proportion of landscape entries giving a location in France including the term 'souvenir', and one or more of 'ruines', 'église', 'abbaye' and 'château', 1790s to 1870s.

The changing nature of artists' and viewers' expectations regarding landscape painting can be read from the levels of use of the term 'vue' [view] in landscape entries. Although in a small minority of cases the entries in which it was used were not specific as to the place depicted, it was predominantly used with those giving a particular location as in 'vue de pont de St-Avertin près Tours' [view of the Saint Avertin bridge, nears Tours] from the Salon of 1808. The definition of 'vue' given by Lévesque and Watelet indicates it would have been understood as suggesting a landscape painting more or less faithful to the location named.⁶⁵ Once domestic locations had grown to be dominant, and as the imaginary classical landscape that featured commonly at the Salon in the 1790s and early nineteenth century was left behind, the use of 'vue' declined rapidly. As Chart 3.13 shows, from featuring in over 50% of landscape entries in the 1830s, the use of 'vue' fell substantially and it featured in fewer than 2% of entries in the 1870s. Artistic practice and viewers' expectations had shifted to the extent that faithfulness to the location named was assumed and no longer needed to be signalled.



Base: 2,207 landscape entries for single oil paintings

Chart 3.13. Proportion of landscape entries for single oil paintings containing the term 'vue', 1790s to 1870s.

⁶⁵ Lévesque and Watelet, (1792), Vol 4., p. 132.

Portrait Entries

Portrait is the generic category in which there was the least change in the form of catalogue entries. From the 1810s onwards, entries were predominantly of the form 'Portrait de' followed by either the name, full or partial initials of the sitter, or the 'trois étoiles' [three stars] device '***'. Entries also often gave the status, title or profession of the individual depicted. Through these honorifics, portrait entries participated in discourses around the self and its relationship to society, alongside the works with which they were associated. The Salon was a highly politicised event and was strongly shaped by the politics of its time. Several art historians have looked at Revolutionary period and nineteenth-century French portraits in terms of the ideas of selfhood they expressed and how those were linked to the circumstances of their production and the prevailing ideologies.⁶⁶

A visual image provides much more information than a catalogue entry and allows a much more nuanced interpretation, but changes in the prevalence of the honorifics used in portrait entries can also be read in relation to those socio-political developments.⁶⁷ My analysis complements and gives support to those other accounts in looking at the whole population of portrait painting rather than, as is the case with the work of other scholars, focusing on a small selection of extant individual works. As I have already discussed, if given in a portrait entry I classified the social status of the sitter into five categories. These were: noble, royal, or imperial; military; governmental including representatives and officials; professional; and, religious. The results from my quantitative analysis of references to social position in portrait entries are presented in Table 3.4.⁶⁸

⁶⁶ See for instance Tony Halliday, *Facing the Public: Portraiture in the Aftermath of the French Revolution*, (Manchester: Manchester University Press, 2000), Amy Freund, *Portraiture and Politics in Revolutionary France*, (University Park: The Pennsylvania State University Press, 2014), Gary Tintertow and Philip Conisbee, (eds.), *Portraits by Ingres: Images of an Epoch*, (New Haven: Yale University Press, 2000), and Heather McPherson, (ed.), *The modern portrait in nineteenth-century France*, (Cambridge: Cambridge University Press, 2001).

⁶⁷ Table 3.1 identifies the regimes in place for each of the Salons included in my analysis.

⁶⁸ As my core sample included only 105 portrait entries for only one Salon from the Second Republic, that of 1849, for the calculations presented in Tables 3.6 and 3.7 I boosted the number of entries by taking an additional 100 portrait entries from the catalogues for each of the Salons of 1848 and 1850.

Social status/Regime	Noble, Royal, or Imperial	Military	Governmental	Professional	Religious	All
Third Republic	9%	4%	2%	1%	2%	18%
Second Empire	10%	2%	3%	3%	3%	21%
Second Republic	3%	1%	2%	1%	2%	9%
July Monarchy	10%	4%	2%	7%	1%	23%
Restoration	28%	6%	4%	7%	2%	47%
First Empire	16%	7%	4%	9%	2%	38%
First Republic	<1%	3%	8%	12%	<1%	24%

Table 3.4. Proportions of portrait entries for single oil paintings containing a reference to the social status of the sitter, by regime.

The First Republic aimed at the radical reconstitution and transformation of the French nation, breaking with the Old Regime and creating new kinds of social relations around the notion of citizenship. As the art historian Lyn Hunt has discussed, members of the new political class identified themselves by their occupation rather than their social rank or lineage.⁶⁹ During the First Republic noble titles were largely absent from the catalogue and 20% of portrait entries described the sitter's professional status as, for instance, a doctor, scientist, or architect, or as a representative or government official.

Noble, royal or imperial titles re-emerged with the creation of the First Empire in 1805 as Napoleon Bonaparte pursued a policy of creating a national elite, including, in 1808, the establishment of a new Imperial Nobility.⁷⁰ Such titles became common during that regime, appearing in 16% of portrait entries. The Restoration monarchies aimed to efface Napoleonic influence and to restore the status and historical significance of the house of Bourbon and of the old nobility. Royal and noble titles appeared in 28% of portrait entries during the Restoration, far more than in any other period.

Most of the period from the 1830s to the 1870s was characterised by the growing political and economic power of the bourgeoisie and of regimes professing to accommodate the various political forces in society. With political ideologies of inclusiveness, signifiers of social status of all kinds became less common during the July Monarchy, Second Empire and Third

⁶⁹ Lynn Hunt, *Politics, Culture and Class in the French Revolution*, (Berkeley: University of California Press, (1984).

⁷⁰ Alexander Grab, *Napoleon and the Transformation of Europe*, (New York: Palgrave, 2003).

Republic than in earlier regimes. They appeared in around 20% of portrait entries during those regimes compared with 38% and 47% of portrait entries during the First Empire and the Restoration respectively. The Second Republic of 1848 to 1851 was a time of great political instability and signifiers of all kinds of social status were far less prevalent in portrait entries during that period than during other regimes.

Historical Entries and Genre Entries

With entries in the genre and historical categories there was little formal change over the period from the 1790s to the 1870s over and above those I have associated with the emergence of titling. As I have discussed, with genre entries this included a significant increase in allusive entries that did not inform the viewer of the manifest subject matter. Otherwise, whether an everyday genre scene such as a mother teaching her child, the return of the hunt, or the harvest, or a particular biblical, mythological, or historical event, entries in these categories continued to be governed by the convention that they give the actors and actions depicted. What remained of importance to artists and to viewers was to indicate and to be informed of the event depicted, and who or what types of people were involved in it.

3.6 Male and Female Artists

The interpretive approach I have set out in this case study can be used to look at questions relating to male and female artists at the Salon. The first is that of whether, on the various measures of linguistic practice I have used, there were any significant differences between women and men in the adoption of titling. To make robust comparisons I have utilised the dataset boosted for women artists. A summary of the main measures I have associated with that process is given in Table 3.5, which shows the very similar paths followed by female and male artists. Both moved away from using their catalogue entries for more than one object at similar rates and to similar extents, as was the case with other measures of individuation. There were substantial declines with entries by both female and male artists in the proportions of their entries providing a reference to the physical object, medium or type of work. The proportions of entries by female and male artists with supportive explicative text both fell substantially.

The small differences in these trends are often explicable, at least in part, in terms of the differences between male and female artists in the types of work and the mix of genres on display at the Salon. In the 1870s paintings in media other than oil such as watercolours and pastels accounted for a higher proportion of paintings on show than in earlier decades. Such

paintings were more likely than oil paintings to have entries for more than one object and to be by women. This is why entries by female artists in the 1850s to 1870s were more likely than those by male artists to be for more than one object. The proportions of entries with supporting explicative text would be expected to be higher for male artists than for female artists as men displayed proportionately more oil paintings and more works with entries in the historical category than women, both of which were more likely to have explicative text than other types of painting or works with entries in other categories.

In summary, there was no material difference between male and female artists in their uses of catalogue entries I have associated with the emergence of titling. In identifying and describing the trends I have associated with that process, generic category is an important dimension whereas the gender of the artist is not. In regard to the adoption of titling, male and female artists behaved in very similar ways. Whilst this is not surprising, my analysis gives further support to my reading of the adoption of titling as a practical and conceptual change that happened across the art world.

	1790s to 1810s	1820s to 1840s	1850s to 1870s
1. Proportion of entries for more than one object.			
Female artists	12%	6%	3%
Male artists	16%	6%	1%
2. Proportion of main entries including the terms 'autre', 'id' or 'idem'.			
Female artists	3%	8%	1%
Male artists	5%	7%	1%
3. Proportion of main entries including a reference to the physical object, medium or type of work.			
Female artists	17%	8%	3%
Male artists	17%	5%	2%
4. Proportion of entries with supporting explicative text			
Female artists	13%	9%	2%
Male artists	16%	12%	6%

Table 3.5. Comparative statistics on language use by female and male artists.

Art historians that have looked at the subject matter of genre painting in the nineteenth-century French art world may not look at gender at the same time, whereas those looking at women artists may not compare male and female artists or look systematically by subject matter.⁷¹ My quantitative approach allows all of those dimensions to be brought together and gives a new way of looking at, and a more detailed understanding of, the inter-relationships of gender, genre and subject matter. In particular, my core dataset has allowed me to investigate the question of how the proportions of oil paintings on display at the Salon made by male and female artists in each generic category changed during the nineteenth century. I will look first at the overall trends, which are shown in Table 3.6.

	1790s to 1810s	1820s to 1840s	1850s to 1870s
Proportion of all still-life entries by female artists	27%	27%	24%
Proportion of all landscape entries by female artists	1%	5%	2%
Proportion of all portrait entries by female artists	23%	16%	13%
Proportion of genre entries by female artists	22%	6%	6%
Proportion of historical entries by female artists	8%	3%	2%

Table 3.6. Comparative statistics on entries in each generic category for single oil paintings by male and female artists, 1790s to 1810s, to 1850s to 1870s.

The painting of still-life subjects was the one area where women had a significant and persisting presence with the oil paintings on display at the Salon. As can be seen from Table 3.6, female artists accounted for 24% to 27% of all entries in the category from the 1790s through to the 1870s. Flowers or fruit were widely considered as the appropriate subjects for women artists. As the critic Albert de la Fizelière maintained in his *Salon de 1850-1851*, painting of 'fleurs et fruit' [flowers and fruit] was 'le seul genre qui convienne aux femmes' [the only genre suited to women].⁷² Still-life subjects to study were also readily available for women artists being taught in the home or at an artist's studios. As I discuss in the Literature Review, in her study of the nineteenth-century American art world, Diana Seave Greenwald argues

⁷¹ Examples of the former are Griffiths, (2013), and Ambrosini, (1995). Examples of the latter are Yeldham, (1994), and Gen Doy, *Women & Visual Culture in 19th Century France: 1800 - 1852*, (London and New York: Leicester University Press, 1998).

⁷² Albert de la Fizelière, *Salon de 1850-1851*, (Paris: Deflorenne, 1851), p. 85.

that still-lives were quicker to produce than paintings in other genres and so were well-suited to women artists who typically had less time to devote to their artistic careers than their male counterparts.⁷³ In contrast, as Linda Nochlin has explored, women were largely excluded from life studies and middle and upper-class women could not travel out of the home unchaperoned, making it difficult for women artists to develop the skills required for history and landscape painting.⁷⁴

We can also look at the still-life subjects indicated by the catalogue entry. For this analysis I used my dataset boosted for female artists to get a better measure of the types of still-life subject painted by women artists. What this shows is that the diversification of still-life subject matter in the oil paintings on display at the Salon was something that happened with both male and female artists, albeit somewhat more strongly with men. During the period from the 1830s to 1870, 43% of still-life paintings displayed by male artists at the Salon represented 'dead nature' compared to 27% by female artists.

My quantitative analysis reveals how strongly gendered the display of landscape paintings in oil at the Salon was throughout the decades from the 1790s to the 1870s.⁷⁵ As can be seen from Table 3.6, there were hardly any such works by female artists on show at the Salon from the 1790s to the 1810s. Although the proportion of landscape entries from female artists rose in the 1820s to the 1840s, as the cultural historian Gen Doy has explored, this was associated with a small number of successful women landscapists.⁷⁶ It fell back in subsequent decades to around 2%. The display of landscapes in oil remained almost completely dominated by male artists.

With oil painting in other generic categories there was a sharp decline in the presence of female artists at the Salon during the nineteenth century. For portraits, the share of all entries by female artists nearly halved over the period I have been examining, falling from 23% in the 1790s to 1810s, to 13% in the 1850s to 1870s. The declines were more extreme with genre and historical entries, with only 6% and 2% respectively of entries in the 1850s to 1870s by female artists. By the 1870s, oil painting of everyday genre scenes and historical subjects had become dominated by male artists to an extent similar to that of landscapes.

⁷³ Greenwald, (2021), p. 98.

⁷⁴ Nochlin, (1988), p. 156.

⁷⁵ For a survey of female artists producing landscapes see Yeldham, (1984), pp. 155 - 156. Yeldham does not compare these artists with male landscape painters.

⁷⁶ Doy, (1998), pp. 85 - 90.

Gen Doy has looked at this question of the reduced presence of women in the fine arts in France in the middle of the nineteenth century.⁷⁷ She argues it can be explained by the increasing pervasiveness of the ideology of the bourgeois family with the role of woman as mother and homemaker. The indications are that in earlier decades of the nineteenth century there may have been less discrimination against female artists, with the situation for women who wanted to pursue a profession in the fine arts slightly easier than later in the century. However, Doy does not look systematically by genre, and her arguments do not explain why the shift was much more pronounced with genre and historical painting than other genres.

To get a better understanding of these changes we can look to the kinds of subjects indicated by the entries in each category.⁷⁸ In the genre category there was a shift from subjects where female artists were well-represented to those where entries from female artists were much less prevalent. Entries indicating an explicitly domestic theme with references to women, children and the home saw a sharp decline. Other types of subject matter such as scenes of military life, or Orientalist scenes from the Middle East or North Africa, became more common during the middle decades of the nineteenth century. With historical entries there was a shift away from subjects drawn from literature, where female artists had a significant presence, and towards patriotic subjects drawn from recent history, which were often battle scenes and, as you might expect, completely dominated by male artists. In contrast, the subject matter of portrait painting did not change in ways that excluded women. Indeed, women artists were nearly as likely as men to have portraits with male subjects on display at the Salon. As with the display of still-life paintings, my quantitative analysis exposes some changes in the nineteenth-century French art world that might merit further investigation. It suggests some possible explanations, such as the growth of Nationalism and Empire, but conventional art historical methods would be required to develop a proper understanding.

⁷⁷ Doy, (1998), pp. 21 - 71.

⁷⁸ In what follows, my ascription of a subject matter is approximate, being based upon keyword searching supplemented by a manual examination (for instance, the keywords for domestic subjects were the nouns 'femme' [woman], 'mère' [mother], and 'enfant' [child], and associated adjectives). In addition, in many cases a subject could not be ascribed with any degree of confidence to genre entries that alluded to the theme of the work. For these reasons, I have refrained from quoting any statistics in the observations that follow.

3.7 Summary and Conclusions

Overview

In this chapter I have set out the first of my three case studies in digital art history. In it I have used descriptive statistical text mining approaches to examine the structure and content of the catalogues of the Paris Salon and Salon criticism. To do this I assembled a dataset from a random sample of entries in the catalogue for Salons covering the decades from the 1790s to the 1870s. For each entry I entered into the dataset a transcription of the main text, together with additional information including the genre of the painting being referenced, the medium used and the number of paintings included under the entry. I boosted the dataset for entries by female artists for my comparisons male and female artists, and assembled a complete set of catalogues in machine-readable format which allowed me to use key-word searching to supplement the analysis of my dataset. I also assembled a small sample of critical writing on the Salon in electronic format.

The readings I give show how the language used by artist and critics was part of practical, conceptual, aesthetic, and institutional change in the nineteenth-century French art world. I have looked at how titling emerged and was adopted as a practice and as a concept to describe that practice. I have suggested that these changes did not happen in isolation but were an integral part of the commercialisation and commodification of the French art world. In relation to genre, my readings show how the language used in still-life entries in the Salon catalogue was involved in the construction and institutionalisation of cultural categories, at how landscape entries signalled aesthetic value in landscape painting, and expressed conceptions of the countryside and its relation to national identity and heritage, and at how the dominant political ideology of each regime was manifested in the honorifics used in portrait entries. In relation to the uses of the language in the catalogue I have examined in this case study, there were no significant differences between men and women artists. We can also see how in some genres, by the middle of the nineteenth-century painting in oil had become heavily dominated by male artists.

Digital Art History

Descriptive text mining of exhibition catalogues and critical writing is an underused approach in digital art history, and the use of sampling in building your dataset is novel within the discipline. These resources and methods allow me to look in aggregate and to take a long-term view on the Paris Salon, examining the decades from the 1790s to the 1870s. They also allow me to investigate the inter-relationships of factors included in my dataset such as those of time, genre, and gender.

The art historical readings I set out in this case study look at the art of the Paris Salon in aesthetic and artistic terms, and also contextually as an embedded practice in which it is viewed from social, economic, institutional and political perspectives. The use of numbers and charts is integral to those readings, for instance in measuring and mapping out how change was manifested.

Throughout this case study I have related my readings to those given by other art historians who have written on the nineteenth-century French art world. My case study shows how the methods of digital art history can be used to develop readings which confirm and extend such scholarship in some ways, and may challenge them in others. My accounts of landscape painting and of portrait painting complement the work of other scholars. I was able to confirm Ruth Yeazell's tentative and impressionistic observation on the emergence of titling in the French art world in the late eighteenth century, and take it forward to show how by the 1870s titling had been widely adopted by artists and critics. My readings challenge the views of those scholar who write about Salon catalogue entries as if they were always titles or on the painting of still-life subjects at the Salon as if they were understood as a unified genre signified in French by the term 'nature morte'. It also suggests questions of art historical interest that may be best addressed through the methods adopted by those scholars, such as the position of still-life painting in the cultural politics of the early 1870s, and that of why the decline in the share of women artists was much greater with oil paintings in certain genres than others.

Standing back from these specifics, this case study shows how digital art historical methods can provide a firmer evidence base for drawing general conclusions than is possible through a reliance on extant paintings or through the non-systematic ways in which art historians may approach archival material. They also allow the art historian to focus in on fine-grained changes which may be of art historical significance.

Methodology

The studies I have examined in the Literature Review are examples of the approach that is adopted by many scholars in the digital humanities. Typically, they draw on one type of data such as a collection of images or a literary corpus. In contrast, the case study I set out in this chapter shows the benefit of utilising several data sources. Through drawing on both Salon catalogues and Salon criticism, I was able to provide a richer and more complete reading of the emergence of titling than would have come from looking at one of them in isolation.

4. THE AUCTION MARKET FOR CONTEMPORARY ART

4.1 Introduction

In his catalogue essay for an exhibition of paintings and drawings by Cy Twombly held at New York's Whitney Museum of American Art in 1979 the cultural theorist Roland Barthes argues that '... in Twombly's titles we must not look for any induction of analogy. If the canvas is called *The Italians* do not look for the Italians anywhere except, precisely, in their name'.¹ In looking at the questions of representation raised by Twombly's use of a proper name in a title, Barthes is one of many scholars for whom their investigation of the relationship between Twombly's titles and the painted or drawn image is an integral part of their interpretive activity. In her 2019 *Reading Cy Twombly: Poetry in Paint* the literary scholar Mary Jacobus often starts a reading by taking her lead from the title.² With Twombly's 1959 series *Poems to the Sea* Jacobus considers in what sense they are poems and what it is to read them as seascapes. In a complex argument she maintains that 'Mallarméan poetics underpin Twombly's practice in *Poems to the Sea*' through the works combining 'the radical contingency of meaningless recurrence (the turn and return of waves) with unfathomable blankness'.³

Even when titles play less of an interpretive role and are used primarily to identify the painting in question, scholars such as Heiner Bastian in his introductory essays for the Catalogue Raisonné of Twombly's paintings focus heavily on his titled works.⁴ However, if we look at the Catalogue Raisonné we find that Twombly did not give titles to 65% of his paintings, a practice that persisted throughout his artistic career. In addition, none of the authors I have surveyed have sought to interpret Twombly's presenting works as 'untitled'. Consciously or unconsciously, many of the scholars who have written on Twombly have focused far more on his titled paintings than on the ones he did not title. The same holds for most of the other artists I look at in this chapter. In the monographs and exhibition catalogues I have reviewed, paintings with specific titles are usually given disproportionate critical attention compared with works that are untitled or have generic titles. A recent monograph by the art historian Karen Kurczynski on Asger Jorn discusses 52 works with specific titles, compared with four that are

¹ Roland Barthes, Heiner Bastian, and Tom Armstrong, *Cy Twombly, paintings and drawings 1954 - 1977*, (New York: The Whitney Museum of American Art, 1979), p. 24.

² Mary Jacobus, *Reading Cy Twombly: Poetry in Paint*, (Princeton: Princeton University Press, 2019).

³ Jacobus, (2019), pp. 83 and 85.

⁴ Heiner Bastian, (ed.), *Cy Twombly: Catalogue Raisonné of the Paintings*, Vols I to VII, (Munich: Schirmer/Mosel, 1997 to 2018).

untitled.⁵ In comparison, Jorn's catalogue Raisonné indicates that 32% of his paintings are untitled or have generic titles such as 'Composition'.⁶ The catalogue for the Sigmar Polke retrospective held at London's Tate Modern in 2008 includes references to 134 paintings by Polke with specific titles and to 37 that are untitled.⁷ Over 60% of Polke's paintings sold at auction have been untitled.

My motivation for the case study I set out in this chapter was to explore this issue of whether the type of title a work has can make a difference to how it is received, but in a different institutional and historical context well-suited to the application of quantitative techniques. In it I look at the auction market for contemporary art and at collectors active in that market over the years from the mid-1980s to the late 2010s. Using statistical methods, I see what auction sales data can tell us about the preferences those collectors may have had between works by contemporary artists presented at auction with specific titles and works presented as untitled or with generic titles such as 'Abstract', 'Number 1' or 'Composition', both of which I will from now on refer to as 'generic' titles.

The critical literature may be an influence on collectors, contributing to a preference for works with specific titles. The ways in which the functions of the title I introduced in Chapter 1 operate with specific and generic titles may also have a similar impact.⁸ Compared with a generic title, a specific title more definitively identifies a work, singling it out from other works by the same artist, and contributes more to the meanings it is given. In doing this the specific title would have more 'seductive' potential than a generic title, attracting potential buyers to the work. Further support for a working hypothesis that collectors have had a preference for works with specific titles comes from the literature on the psychology of art reception. In several empirical studies in which participants are asked to rate paintings presented with and without titles, researchers have shown that the inclusion of a title with a painting can positively affect the viewer's understanding or liking of the work of art.⁹

If collectors had a preference for paintings or sculptures with specific titles, they would be prepared to pay more at auction for such a work than for a comparable work by the same artist

⁵ Karen Kurczynski, *The Art and Politics of Asger Jorn: The Avant-Garde Won't Give Up*, (Ashgate: London, 2014).

⁶ Guy Atkins, (ed.), *Asger Jorn: Catalogue Raisonné*, Vols 1 to 4, (London: Lund Humphries, 1968 to 1980).

⁷ Kathy Halbreich, (ed.), *Alibis: Sigmar Polke 1962-2010*, (London: Tate Publishing, 2014).

⁸ Genette, (1975), and Besa Camprubi, (2003).

⁹ For a recent study and references to earlier work see John W. Mullenix and Julien Robinet, 'Art Expertise and the Processing of Titled Abstract Art', *Perception*, 47(4), (2018), pp. 359 - 378.

presented with a generic title. However, it would be wrong to attempt to see if there is such a premium paid at auction by comparing the simple statistic of the average sales price for works with each type of title as that is to ignore the differences there are between works of art and the date and circumstances of their sale. To develop an understanding of the preferences collectors may have had for different types of title I turned to an explanatory approach that is well-established with cultural economists for the analysis of the auction market and takes account of the interactions of all these factors, that of 'hedonic' modelling.

Hedonic models were given that name as they were originally developed from the economic theory that non-standardised goods can be modelled as bundles of characteristics, each of which is a source of utility or pleasure and is valued separately by the consumer.¹⁰ In applying this approach to auction sales of works of art, the price is related to characteristics of the work of art such as its size, medium, genre, provenance, and the colours used, and to characteristics of the artist such as their life status, age, gender and level of critical or art institutional recognition. Hedonic pricing models may also look at potential determinants of the price not directly related to the artwork or artist such as the auction house and the date and location of the sale. The statistical technique of linear regression is then used to give an understanding of which of these characteristics make a statistically significant contribution to the overall price and to estimate the size of that contribution.

An early application of hedonic regression modelling to auction sales of paintings was by the economist Robert C. Anderson.¹¹ In his 1974 study, the model he developed gave an understanding of the price achieved at auction with 1,500 sales of paintings over the years from 1690 to 1960 in terms of an underlying rate of appreciation, the size of the painting and a measure of the artist's reputation. As I will now discuss, since then cultural economists have used regression modelling to look at a range of questions around the influences of the price achieved at auction with sales of works of art. Scholars may investigate a specific factor which impacts upon the price, using regression models to control for the influence of other factors. Others have looked broadly at the price determinants or financial returns across the fine art market or in a particular segment. Often scholars compare their results on the influence of factors such as the size, medium and auction house with the results of other studies. In recent

¹⁰ For an introduction to hedonic modelling see Ben Sopranzetti, 'Hedonic Regression Models' in Cheng-Few Lee and John C. Lee, (eds.), *Handbook of Financial Econometrics and Statistics*, (New York: Springer, 2014), pp. 2,119 - 2,134.

¹¹ Robert C. Anderson, 'Painting as an Investment', *Economic Enquiry*, 12/1, (1974), pp. 13 - 26.

years cultural economists have taken advantage of the availability of auction data in digital format to look at very large volumes of auction sales.

Amongst studies looking at a particular determinant of auction prices, David Galenson studied the relationship between the auction value of a painting and the artist's age at execution.¹² His study drew on data for around 4,500 auction sales covering the years from 1980 to 1996 of paintings by 42 American or America-based contemporary artists born before 1940. Douglas J. Hodgson conducted an analysis of the age-price relationship with 26,955 auction sales over the years from 1968 to 2010 of paintings by 120 Canadian artists covering the entire history of Canadian art.¹³ Looking at over 430,000 sales at auction over the years from 1980 to 2005, Heinrich W. Ursprung and Christian Wiermann analysed how an artist's death influences the market prices for their art.¹⁴ Alan Beggs and Kathryn Graddy looked at how previous sales of Impressionist, Modern and Contemporary paintings influence the price paid at auction when a work returns to market.¹⁵ Helen Higgs and Jon Forster investigated the preference amongst purchasers of Australian art for paintings of different sizes.¹⁶ Their dataset was around 52,000 sales of works by 70 Australian artists at Australian auction houses over the years 1986 to 2009. Drawing on auction sales data for over 65,000 sales of sculptures in 43 different countries over the years from 1985 to 2013, Rustam Vosilov considered the question of whether there was a home bias amongst collectors.¹⁷ Looking at around 800 sales over the years from 1955 to 2015, Kim Oosterlinck and Anne-Sophie Radermecker investigated the question of whether art market participants value provisional names ('The Master of ...') with paintings by Flemish Old Masters.¹⁸ Elena Stepanova studied the questions of whether the particular colours used in a painting, and the colour diversity of a composition can impact the price paid at auction. Her dataset was 127 sales of paintings by Picasso over the years from

¹² Galenson, D., 'The Careers of Modern Artists: Evidence from Auctions of Contemporary Paintings', NBER Working Paper 6331, (December, 1997).

¹³ Douglas J. Hodgson, 'Age-price profiles for Canadian painters at auction', *Journal of Cultural Economics*, 35, (2011), pp. 287 - 308.

¹⁴ Heinrich W. Ursprung, H., W. and Christian Wiermann, 'Reputation, price, and death: an empirical analysis of art price formation', CESifo Working Paper, No. 2237, (2008).

¹⁵ Alan Beggs and Kathryn Graddy, 'Anchoring Effects: Evidence from Art Auctions', *The American Economic Review*, 99/3, (2009).

¹⁶ Helen Higgs and John Forster, 'The Auction Market for Artworks and their Physical Dimensions including the Golden Ratio: Australia - 1986 - 2009', SSRN Research Paper, (June, 2011).

¹⁷ Rustam Volikov, 'Art Auction prices: Home Bias, Familiarity and Patriotism', SSRN Research Paper, August, (2015).

¹⁸ Kim Oosterlinck and Anne-Sophie Radermecker, "'The Master of ...': creating names for art history and the art market", *Journal of Cultural Economics*, August, (2018).

1998 to 2016, and 371 sales of paintings by Color Field Abstract Expressionists over the same period.

Scholars have looked at financial returns and price determinants across a wide range of market sectors, and at the fine art market as a whole. Madeleine de la Barre, Sophie Ducclo and Victor Ginsburgh used regression modelling to compare the returns on auction sales of Impressionist, Modern and Contemporary paintings by European artists.¹⁹ In their study they drew on auction sales data for 82 artists associated with those movements, covering the years from 1962 to 1992. Richard J. Agnello and Renée K. Pierce studied price determinants including genre effects and financial returns with around 15,000 auction sales of works by 66 American artists over the years 1971 to 1992.²⁰ In a similar study, Luc Renneboog and Tom Van Houtte looked at sales of paintings by Belgian artists associated with movements from Realism to Surrealism.²¹ Their dataset consisted of over 10,500 sales covering the years from 1970 to 1997. Prior to their study of size, Helen Higgs and Andrew Worthington carried out a general investigation of the market for Australian artists, looking at around 37,000 sales of works by 60 Australian artists over the years 1973 to 2003.²² Luc Renneboog and Christophe Spanjaers utilised a dataset of 1.1 million auction sales of works by over 10,000 artists held worldwide over the years from 1960 to 2007 to look at returns across the fine art market and at a range of characteristics which might influence the price such as the presence of a signature, the subject and the auction house and location.²³ Mathieu Aubry, Roman Kräussl, Gustavo Manso, and Christophe Spaenjers have used regression modelling as a benchmark against which to measure the performance of convolutional neural networks in predicting the price of over 1.1 million works of art sold at auction over the years 2008 to 2015.²⁴ Another

¹⁹ Madeleine de la Barre, Sophie Ducclo, and Victor Ginsburgh, 'Returns of Impressionist, Modern and Contemporary European Paintings, 1962 - 1991', *Annales d'Économie et de Statistique*, 35, (Jul. - Sep. 1994), pp. 143 - 181.

²⁰ Richard J. Agnello, and Renée K. Pierce, 'Financial Returns, Price Determinants, and Genre Effects in American Art Investment', *Journal of Cultural Economics*, 20/4, (1996), pp. 359 - 383.

²¹ Luc Renneboog and Tom Van Houtte, 'The Monetary Appreciation of Paintings: from Realism to Magritte', *Cambridge Journal of Economics*, 26/3, (2002), pp. 331 - 358.

²² Helen Higgs and Andrew Worthington, 'Financial Returns and Price Determinants in the Australian Art Market, 1973 - 2003', *The Economic Record*, 81/253, (June, 2005), pp. 113 - 123.

²³ Luc Renneboog and Christophe Spaenjers, 'Buying Beauty: On Prices and Returns in the Art market', *Management Science*, 59/1, (2013), pp. 36 - 53.

²⁴ Mathieu Aubry, Roman Kräussel, Gustavo Manso, and Christophe Spaenjers, 'Machines and Masterpieces: Predicting Prices in the Auction Market', March, (2019). HEC Paris Research Paper No. FIN-2019-1332.

example of the application of machine learning techniques for the prediction of auction prices is the work of Jason Bailey.²⁵

In my case, the regression model I developed was one which included the type of title as one characteristic of an artwork along with other factors relating to the work itself, to the artist and to the circumstances of the sale. The main difference between my modelling approach and that of most of the authors I have surveyed is that I model each artist separately rather than looking at the market as a whole or a market sector. The primary reason is that it is potentially of greater art historical interest to understand the preferences collectors may have had for otherwise comparable works by the same artist with different types of title, than to look at averages across the market. This approach also allowed me to look at similarities and differences between the preferences of collectors of different artists. David Galenson is the other scholar who has developed models for individual artists, as he wanted to compare the ways in which the artist's age at execution of the work of art impacted upon the sales price.

To develop my model, I collected data on auction market sales for works by contemporary artists held over the years from 1984 to 2019 – where a 'contemporary' artist is someone whose works have been sold at auctions advertised by the major auction houses as being of 'contemporary' art. I looked for artists with a significant international presence in the auction market whose sales of paintings or sculptures consisted of some works presented at auction with generic titles, and others with specific titles.

Altogether, twelve artists met my search criteria. All have been and remain among the top selling contemporary artists at auction. All four born post-war feature regularly in the top 25 of the art market information provider Artprice.com's annual list of best-selling contemporary artists, and the other eight have sales that would rank them in the top 75.²⁶ The next section gives an introduction to the auction market, details the process I followed to identify the twelve artists I have modelled, and gives an overview of their titling practices as given in the titles of their works as presented at auction and, where available for me to consult, their Catalogues Raisonnés.

After having introduced the artists that I model, I give a critical discussion of the sales data used in my analysis and an introduction to regression modelling. My aim is to give the reader

²⁵ Jason Bailey, 'Can Machine Learning Predict the Price of Art at Auction?', *Harvard Data Science Review*, 2/2, (2020).

²⁶ *The Contemporary Art Market Report 2019*, (Artprice, 2019). Artprice defines a contemporary artist as someone born after 1945.

an appreciation of how regression analysis works, and of the data, methodological and interpretive issues involved in its use, so it is not treated as a 'black box'. I then move on to give my reading of the results of the modelling, looking at what my models say about collectors' preferences and about the structure of the auction market. I begin by considering features of works of art such as their size and the impact of the location of the sale on the price achieved at auction. I have started with these results as they help in understanding other aspects of my model. I end my reading by looking at the type of title a work is presented with at auction. Although the number of artists I have looked at is small, in my models there are some clear patterns in relation to collectors' preferences across them. In addition to giving a reading of my model, I also look at how my results compare with studies of the auction market by cultural economists, and how they relate to the broader conclusions drawn in that literature.

A regression model can give an explanation of the price achieved at auction though relating it to various characteristics of the work, of the artist and of the circumstances of the sale. What it does not do is explain why some characteristics are a significant influence on the price whereas others are not. Cultural economists may stand back from their models and propose reasons why that is so that are not supported by the data. As I discuss the results of my modelling I will also venture some speculative and qualitative answers to these questions.

In the concluding section to this chapter I summarise the reading presented in this case study and look at the answers it gives to the broad questions which have shaped my work. I also reflect upon the methodological issues raised in using regression analysis in an art historical context.

The art historical value of my work is not limited to that of giving an understanding of the influence of the type of title on the price achieved at auction. For all the characteristics of art works I have looked at, the application of regression analysis to auction sales data has allowed me to bring new knowledge into art history. With traditional art historical methods, it is not possible to develop the kind of disaggregated understanding of collectors' preferences that I present in this chapter. Although the work of some cultural economists such as Cynthia White and Harrison White's study of institutional change in the late nineteenth-century French art world has been influential with art historians, much less attention has been paid to the type of work I have drawn on in this case study.²⁷ A review of the citations for the papers I have surveyed reveals only one by an art historian.

²⁷ White and White, (1965).

My work on collectors' preferences for different types of title with particular artists is new within cultural economics. In other areas such as the influence on the sales price of the size of the painting and of the life stage of the artist when executed, my work complements and re-contextualises previous scholarship, showing how collectors' preferences and market structures have changed.

As with the other chapters of my thesis I have looked to minimise the amount of technical detail in the main sections. For the interested reader, a brief introduction to regression modelling and the details of the steps I went through to collect my data and develop the models I present and interpret in this chapter is given in Appendix B.

4.2 The Auction Market, Artists, Works of Art, and Titles

A specific auction market for contemporary art had its beginnings in the United States in the late 1960s and early 1970s.²⁸ At that time Sotheby's staged sales advertised as being of 'contemporary' art once a year in New York. In 1973 it established a contemporary art department, and since then has held contemporary art sales twice a year in New York and London, to be followed by Christie's in 1974. Originally heavily dominated by New York and then by that city with London, the auction market for contemporary art has expanded considerably since those early days. Through the 1980s auction houses in a number of countries were holding sales of contemporary art, although the borders between what was sold as contemporary art or as modern or post-war art were quite fluid. By the early 1990s, in the words of the art historian Terry Smith, '... sage observers could see that contemporary art had developed its distinctive structures...'.²⁹ It was recognised in critical and academic literature and a canon had been established. With globalisation and financial deregulation bringing new collectors into the market from the early 1990s onwards, auction sales of contemporary art were established in many countries including several in Asia. China, in particular, has become a major international location for sales of contemporary art, with a strong domestic market. Since the early 2000s Chinese auction houses have featured regularly in the top ten for annual sales of contemporary art by turnover in the art market website Artprice.com's annual rankings.³⁰ Since 2017 Hong Kong has had a contemporary art week held twice year where all of the main auction house stage sales. Contemporary art has

²⁸ For histories of the auction market for contemporary art see Terry Smith, *What is Contemporary Art?*, (Chicago: Chicago University Press, 2009), pp. 117 - 132, and Noah Horowitz, *Art of the Deal: Contemporary Art in a Global Financial Market*, (Princeton and Oxford: Princeton University Press, 2011), pp. 3 - 21.

²⁹ Smith, (2009), p. 128.

³⁰ [Artprice. \(2019\).](#)

also become the most important fine art category for the major auction houses. In 2007 sales of contemporary art by Sotheby's and Christie's exceeded their sales of Impressionist and Modern art for the first time. In 2018 these two auction houses grossed around £2.4 billion at sales of contemporary art, compared to £1.8 billion at sales of old masters, Impressionist and Modern art combined.

To identify contemporary artists to model I searched the auction sales database provided by Artprice.com.³¹ Sotheby's and Christie's define their contemporary art sales as encompassing artists and schools or movements from early Abstract Expressionism to the current day. However, in practice artists who were active in the 1930s or early 1940s as well as in the Post-War period are also included. I followed these definitions with the other auction houses included in my search. As I wanted to develop separate models for each artist and to consider the impact of the location of the auction on the sales price, I looked for artists with an international presence in the auction market, including a significant number of auction sales in several countries of 200 or more in total, and an average sales price of over \$50,000. I restricted the sales to one of either paintings or sculptures for each artist as this allowed me to develop models that are simpler and easier to understand, and permits comparisons to be made between artists.³² Each of the artists was chosen to have a mix of sales of works presented as untitled or having short titles including the generic words 'abstract', 'number', or 'composition', and works with specific titles. I restricted my list of artists further by removing artists where all of the works with generic or specific titles were executed in a short period of their career or where the thumbnail images of the paintings provided by Artprice.com suggested that works with different kinds of title were visually distinct. There were also several artists for whom I was unable to develop a satisfactory model. Taken together, these criteria narrowed down the number of artists I could investigate and applying them lead me to the twelve artists I discuss in this chapter.

For eleven of these artists I have looked at sales of their paintings. Alexander Calder is the exception, and it is sales of his sculptures I have modelled. For twelve of the artists I have used the auction sales data by Artprice.com. For one artist, Gerhard Richter, his official website also provides auction sales data, and I have used that in my modelling as it is a more

³¹ The Artprice database can be found at <https://artprice.com>.

³² In doing this I have relied upon the categorisation given by Artprice.com, which groups works into sculptures, paintings, drawings, and prints or editions.

consistent data source than Artprice.com which consolidates its data from a large number of auction houses.³³

The collection of artists I have assembled shows how heavily biased the contemporary art auction market has been towards male artists, which was most pronounced in the twentieth century. Few women have had a successful and extended career at the top end of the auction market over the period I have looked at. Only one of the twelve artists who passed my selection criteria was a woman, and, as will be seen, most of the auction sales of her paintings have been in the current century. One other female artist, Yasoï Kusama, met most of the selection criteria I used, but she was one of the artists for whom I was unable to build a satisfactory model. The prices paid for her paintings have been very volatile, varying substantially at different times and for works created at different stages of her career.

Five of the twelve artists I have modelled are American. Alexander Calder (1898 - 1976) worked primarily as a sculptor and also produced paintings, prints and jewellery. The sculptures were works in wire and what Calder called 'mobiles' and 'stabiles'.³⁴ Mobiles are suspended or standing sculptures that move mechanically or with the flow of air around them, and stabiles are stationary sculptures. In my work, as I explain in Appendix B, I have looked only at sculptures identified as mobiles in the auction sales data. The most common types of specific title with Calder's mobiles suggest animal or other natural world associations, or provide formal descriptions, sometimes in combination. Occasionally the specific titles are humorous or whimsical. Mobiles may also be presented at auction as untitled, or in some cases with other generic titles such as 'Standing Mobile'. Works with generic titles have made up around 40% of the auction sales of Calder's mobiles.

The second American artist I have looked at is the painter Sam Francis (1923 - 1994). The *Catalogue Raisonné* of Sam Francis' paintings on canvas and board lists 1,855 works produced over the years from 1946 to 1994, the majority of which are listed as untitled.³⁵ Most commonly, when presented at auction these paintings will also be listed as untitled, although for a significant number of sales the auction house will give the Sam Francis Foundation Archive number as the title or as a bracketed sub-title (the paintings in the *Catalogue Raisonné* are numbered from SFF.1 to SFF.1,855). Francis occasionally used other types of generic title

³³Gerhard Richter's official website can be found at <http://www.gerhardrichter.com/>.

³⁴ Alexander Calder recounts the stories over his adoption of the terms 'mobile' and 'stabile' for his sculptures in his autobiography. Alexander Calder, *Calder: An Autobiography with Pictures* (New York: Pantheon Books, 1966).

³⁵ Debra Burchett-Lere, (ed.), *Sam Francis: Catalogue Raisonné of Canvas and Panel Paintings, 1946 - 1994*, (Berkeley: University of California Press, 2011).

such as 'Composition'. Together, works with generic titles have made up nearly 70% of the auction sales of Francis' paintings. The remaining 30% are presented with specific titles which are mainly formal or descriptive, or may be anecdotal, whimsical or refer to particular locations.

Joan Mitchell (1925 - 1992) worked primarily as a painter. Her paintings have appeared at auction over the whole of the period I have modelled, but sales volumes have been much higher in the last two decades, with over 40% in the 2010s. Around 48% of auction sales of Mitchell's have been of paintings presented with generic titles, primarily 'Untitled' with a small number titled 'Composition'. The remaining 52% of sales are of paintings with specific titles, often alluding to a personal memory or emotion associated to a thing or a place that Mitchell connected with the painting.³⁶ From 1955 Mitchell split her time between the United States and France, and lived from 1968 in a town on the outskirts of Paris. She is the one artist of the twelve I have modelled with a large volume of auction sales in France.

Over the course of his career Cy Twombly (1928 - 2011) produced over 500 paintings, which are presented in the Catalogue Raisonné of his paintings with titles as agreed with the artist.³⁷ The majority of the paintings are listed as 'Untitled', and very occasionally Twombly added a sub-title in brackets, a practice which became much more common in the years from 2008 to 2011. The remaining paintings have specific titles, most commonly alluding to mythological or classical subjects, with others including references to Renaissance or art historical subjects, or to locations or subjects taken from the natural world. Sometimes Twombly reworked past subjects using the same or similar titles, and added a number to indicate works that were part of a cycle or series. Paintings with specific titles have represented 35% of the auction sales of Twombly's paintings.

The final American artist is Christopher Wool, (1955 -), whose paintings have appeared at auction since the early 1990s. Around 20% of the sales of Wool's paintings at auction have been of works with specific titles. These titles may single out some text from a stencilled word painting, or evoke cross-cultural references. The remaining 80% are for works presented as untitled or with the number given to the painting by Wool used as the title.

Five of the other artists I have modelled are German. For all of them, their paintings are most often presented at auction with German titles, sometimes with an English translation in brackets, and others are presented with titles in English or other languages. Martin

³⁶ Linda Nochlin, 'Joan Mitchell a Rage to Paint', in Jane Livingston, (ed.), *The Paintings of Joan Mitchell*, (Berkeley: University of California Press, 2002), p. 58.

³⁷ Heiner Bastian, (ed.), (1997 to 2018).

Kippenberger lived from 1953 to 1997, and his paintings have appeared in significant numbers at auction from 1999 onwards. The specific titles he gave to his paintings have been read as verbal jokes or clues to the complex visual puzzles which play out in his paintings.³⁸ Works with specific titles have accounted for 67% of the sales of Kippenberger's paintings. The remaining 33% were presented at auction as untitled.

The second German artist I have modelled is Albert Oehlen (1954 -). Oehlen's paintings have appeared at auction since the middle of the 1980s, but sales volumes are concentrated in the years since the early 2000s. Works with specific titles have made up 65% of auction sales. These titles are one of the many elements of painting that Oehlen has been read as playing off against each other in his artistic practice.³⁹ The remaining 35% have been of paintings presented at auction as untitled.

A. R. Penck was the moniker under which Ralph Winkler (1939 - 2017) produced most of his art. Around 30% of the auctions sales of paintings by A. R. Penck have been of works presented as untitled. The remaining 70% were of paintings with specific titles, often alluding to his interests in power systems and means of resistance and change.⁴⁰

The specific titles of paintings by Sigmar Polke (1941 - 2010) have been read as working allusively, often with irony or humour, to engage with his strategies of social, political and aesthetic critique.⁴¹ His other paintings are presented at auction as untitled or as untitled with a sub-title in brackets. Works with specific titles have accounted for around 40% of the sales of paintings by Polke.

The final German artist is Gerhard Richter (1932 -). Since the middle of the 1970s Richter has produced a body of nearly 2,000 paintings that are classified as abstracts on his official website.⁴² In titling those works he has followed three approaches. The most common approach is to title the work 'Abstraktes Bild' [Abstract Painting]. A small number of abstract works are catalogued as 'Ohne Titel' [Untitled] or have a date recorded as their title. The remainder of Richter's abstracts I have included in my modelling comprise works with specific

³⁸ Gregory Williams, 'Jokes Interrupted: Martin Kippenberger's Receding Punch Line', in Doris Krystof and Jessica Morgan, (eds.), *Martin Kippenberger*, (London: Tate, 2006), pp. 39 - 48.

³⁹ Albert Oehlen, 'Born to be Late', in *Selections from the Collection of Pierre Huber*, (New York: Christie's, 2007). Exhibition Catalogue. Albert Oehlen, 'Treppe', in *Contemporary Art Day Auction*, (Sotheby's, 2019). Exhibition Catalogue.

⁴⁰ Lena Fritsch, *A. R. Penck: I Think in Pictures*, (Oxford: Ashmolean Museum, 2019).

⁴¹ Mark Godfrey, 'From Moderne Kunst to Entartete Kunst: Polke and Abstraction', in Kathy Halbrich, (2014), pp. 118 - 143.

⁴² The statistics presented on Gerhard Richter's abstracts in this section are correct as of June 2019.

titles, which are sometimes thematically related and grouped together within his catalogue, and small groups of works that have been given the same title.⁴³ In my modelling I included as generic titles 'Abstraktes Bild', 'Ohne Titel' or a date, and categorised all other titles as specific. Works with specific titles have accounted for around 20% of the auction sales of Richter's abstracts.

The last two artists I have investigated are the Japanese painter Yoshimoto Nara (1959 -) and the Danish artist Asger Jorn (1914 - 1973). The youngest artist I have modelled, Nara is the only artist with a large volume of sales in Japan, Hong Kong or mainland China. Nara typically titles his paintings in English, and their specific titles often work to reinforce the tension in his paintings between the apparent naivety of the characters or animals he paints and their actions or gestures.⁴⁴ Nara's paintings have appeared at auction since the early 2000s. Works with specific titles have made up nearly 80% of those auction sales. The remaining 20% of auction sales were of paintings presented as untitled.

The Catalogue Raisonné of Asger Jorn's paintings lists over 1,800 works executed over the years from 1940 to 1973 and which, as discussed by the editor Guy Atkins, indicate his titling practice went through several phases.⁴⁵ In the period from 1940 to 1952 Jorn used Danish almost exclusively for his specific titles, which often included imaginary and real placenames. In later years Jorn titled most often in French, English, German and Italian as well as Danish, with specific titles that were often ironic or humorous involving inventive wordplay, and could allude to high and popular culture. Together, works with specific titles have represented 69% of the auction sales of his paintings. Jorn did not give titles to many of his paintings, and paintings appearing at auction with other generic titles such as 'Figure' or 'Composition' may represent later retitlings by a dealer or owner of an untitled work. Auction sales of Jorn's paintings have been predominantly in Europe and he is the only artist I have modelled who has not had a significant presence at auctions held in the United States.

⁴³ Richter's website lists four long series of abstract paintings, with all works in each series given the same title. These are the 115 works titled 'Grün-Blau-Rot' [Green-Blue-Red] from 1993, the 110 works titled 'Fuji' from 1996, the 64 works titled 'Miniatüren' [miniatures], also from 1996, and the 64 works titled 'Schwarz, Rot, Gold' [Black, Red, Gold] from 1998. Although unique paintings the works in each series are visually and materially very similar and, as well as appearing in Richter's Catalogue Raisonné, they also appear in the Catalogue Raisonné of his Editions. Editions typically command a much lower price at auction than unique works of art by the same artist, and so I excluded them from my analysis.

⁴⁴ Mika Yoshitake and Michael Govan, *Yoshimoto Nara*, (London: Prestel, 2020).

⁴⁵ Atkins, (1968 - 1980), Vol. 1, pp. 119 - 120, Vol. 2, pp. 143 - 147 and Vol. 3, p. 24.

4.3 Collectors' Preferences for Contemporary Art

Data

The auction sales data provided by Artprice.com and on Gerhard Richter's official website includes the title of the work, the year of its creation, the medium such as oil or acrylic, the support such as canvas or board, and its dimensions.⁴⁶ The date and location of the sale, the auction house, and the sale price or a flag if the lot was not sold are also provided.⁴⁷ In my modelling I only included sold lots, and to allow for comparisons across countries and at different times I looked at prices in US Dollars and in real terms, that is adjusting for inflation.⁴⁸

To automate my sampling of the Artprice.com portal I used scripts written in the open-source language Python. Python scripts were also used to help clean and consolidate the sales data and to partially automate the model development process. The sales data for all the artists I have modelled was consolidated into an Excel workbook.

The period I have investigated was initially determined by the availability of data - the Artprice.com sales data goes back to 1984, and the data I used goes through to June 2019. The data provided on the Gerhard Richter website for his abstracts goes back to 1987. In the event this turned out to be a suitable time frame for my analysis. A shorter time frame would have significantly reduced the number of artists I could model as it would have included fewer auction sales. The risk of using a long time frame is that collectors' preferences may have changed substantially over that period. However, as I will come on to discuss, the results of my modelling indicate that does not appear to have been the case over the period I have used.

To prepare the auction sales data for modelling I had to categorise those characteristics that are not numerical. As I have already discussed, I categorised titles as generic or specific. In many cases the Artprice.com data does not include the support, and so I did not include that characteristic in my modelling. For the medium, I took paintings in oil, or in oil and other media, as one category, and all other media as another. In my modelling I wanted to look both at how the two dominant auction houses, Sotheby's and Christie's, compared to others and also at

⁴⁶ I have used Richter's website in my analysis rather than Artprice.com - the source for the other artists investigated in this section - as Richter's website records more sales and the data is more consistent than the Artprice.com data, which is collected and consolidated from a large number of auction houses.

⁴⁷ The Artprice.com sales data gives the hammer price, whereas in most cases Richter's official website gives the price including the buyer's premium, a charge on the lot paid to the auction house, and I excluded other sales from my model. Buyer's premiums are typically in the range 10% to 25% of the hammer price, with rates varying by location, and changing over time and with the purchase value. This should be borne in mind when comparing average sales prices for paintings by Richter and other artists.

⁴⁸ June 2019, the most recent month for which I have obtained auction sales data, is the baseline month.

how the location of the sale might have influenced the prices achieved at auction. I therefore grouped together Sotheby's and Christie's as one category and all other auction houses as another. For each artist, I included the locations with a significant number of sales as separate categories, grouping the remainder together as another. For all artists bar Joan Mitchell and Christopher Wool the United Kingdom featured as a category, and for all bar Asger Jorn the United States. As sales in the United States was a category in my model for all artists, the native country category did not include American artists. For Yoshimoto Nara, Hong Kong also featured, as did France for Joan Mitchell, the country in which she settled in 1968.

With the numerical data, I combined the dimensions of the paintings into one measure of area and for Alexander Calder's mobiles I took the longest dimension reported in the sales data. I took the year of creation and converted it into the age of the artist when the work was executed and chose a form for the model allowing me to look at how prices can vary for works created at different times of an artist's career. For each sale, I included the number of months of the sale date from December 1983, which allows an underlying rate of appreciation to be determined for the auction value of paintings by each artist.

My model includes an allowance for a boom in the contemporary art market in the late 1980s which affected all the artists I have modelled with sales during that period. It also allows for persisting downturns in the 1990s for the auction prices with paintings by Sam Francis, Asger Jorn and A. R. Penck, and for a recent upturn for Penck. For the other nine artists, other than the boom in the late 1990s where relevant, there has been consistent underlying growth in their sales prices at auction.

Some cleaning was required to correct for inconsistencies in the Artprice.com data and I excluded sales where the size or date of sale was not provided. Finally, some sales were excluded as part of the model development process including those where the sales price far exceeded that for any other sales by the artist, and particular groups of paintings whose sales prices were consistently over- or under-predicted by my models. Across the twelve artists, there are a total of 6,197 auction sales included in my modelling.

Table 4.1 gives the average sales prices achieved at auction for the twelve artists I have looked at and for each of the categories used in my modelling. The averages are restricted to the sales I have included in my modelling, with the total number of such sales given in the first row of the table. Looking first at the differences between works sold with specific titles and those with generic titles, in most cases the former sell for much more on average than the latter, although that is reversed for Martin Kippenberger and Cy Twombly. For all artists, sales

at Sotheby's or Christie's result in a significantly higher average sales price than sales at other auction houses. Auctions in the artist's native country deliver much lower sales prices on average than those in the United States or the United Kingdom. There is no pattern to whether paintings in oil sell for more or less on average than works in other media. However, as I have already discussed, it would be wrong to conclude that collectors of Kippenberger have placed a premium on generic titles over specific titles or that sales at Sotheby's and Christie's delivered better results for consignors than sales at another auction house. These factors cannot be considered in isolation as that ignores the relationships between them and with other factors that may determine the price such as the size of the work, the medium, and the date and location of the sale. The model I have developed allows us to get an understanding of the inter-relationships between all these factors and of how they affect the sales price.

	Alexander Calder	Sam Francis	Asger Jorn	Martin Kippenberger
Number of sales	853	489	868	311
Oil	n/a	\$947,878	\$111,553	\$690,082
Other medium	n/a	\$149,862	\$36,186	\$231,820
Sale at Sotheby's or Christie's	\$976,315	\$461,743	\$107,825	\$613,364
Sale at another auction house	\$871,657	\$133,133	\$93,436	\$193,697
Sale in the United States	\$1,018,565	\$521,982	n/a	\$630,958
Sale in the United Kingdom	\$869,191	\$219,454	\$124,395	\$558,833
Sale in Hong Kong	n/a	n/a	n/a	n/a
Sale in France	n/a	n/a	n/a	n/a
Sale in native country	n/a	n/a	\$100,233	\$47,640
Sale in other country	\$631,837	\$154,181	\$81,749	\$184,708
Generic title	\$745,067	\$178,264	\$56,643	\$743,390
Specific title	\$1,098,192	\$781,849	\$119,039	\$293,032
All sales	\$969,444	\$381,078	\$99,918	\$439,290

Table 4.1. Auction Sales Statistics for Artists included in my modelling.

	Joan Mitchell	Yoshimoto Nara	Albert Oehlen	A. R. Penck
Number of sales	326	483	233	527
Oil	n/a	\$221,139	\$483,122	\$68,525
Other medium	n/a	\$338,765	\$417,507	\$50,565
Sale at Sotheby's or Christie's	\$1,448,948	\$417,334	\$561,366	\$72,778
Sale at another auction house	\$937,542	\$187,739	\$237,799	\$39,124
Sale in the United States	\$1,440,894	\$332,650	\$430,298	\$73,074
Sale in the United Kingdom	n/a	\$303,981	\$575,050	\$76,839
Sale in Hong Kong	n/a	\$573,967	n/a	n/a
Sale in France	\$1,103,208	n/a	n/a	n/a
Sale in native country	n/a	\$56,131	n/a	\$31,057
Sale in other country	n/a	\$175,759	\$65,447	\$47,676
Generic title	\$1,024,184	\$130,621	\$416,512	\$43,178
Specific title	\$1,709,244	\$378,535	\$478,903	\$60,501
All sales	\$1,390,184	\$324,640	\$457,214	\$55,472

Table 4.1. Continued.

	Sigmar Polke	Gerhard Richter	Cy Twombly	Christopher Wool
Number of sales	413	467	232	364
Oil	\$517,753	n/a	\$2,094,308	n/a
Other medium	\$731,739	n/a	\$4,939,122	n/a
Sale at Sotheby's or Christie's	\$910,316	\$3,627,013	\$3,874,220	\$1,033,731
Sale at another auction house	\$168,828	\$1,285,289	\$2,059,717	\$639,779
Sale in the United States	\$854,405	\$4,444,286	\$4,696,690	\$1,045,992
Sale in the United Kingdom	\$887,531	\$2,756,993	\$2,364,575	\$421,551
Sale in France	n/a	n/a	n/a	n/a
Sale in Hong Kong	n/a	n/a	n/a	n/a
Sale in native country	\$114,049	\$481,354	n/a	n/a
Sale in other country	\$237,509	\$1,294,795	\$421,805	\$365,778
Generic title	\$354,047	\$2,770,294	\$4,564,146	\$726,367
Specific title	\$1,200,625	\$5,780,921	\$2,481,138	\$1,461,547
All sales	\$714,249	\$3,331,160	\$3,704,351	\$869,775

Table 4.1. Continued.

Modelling and Interpretation

As I mentioned in the introduction, in a linear regression model for auction sales the price achieved at auction is related to a number of characteristics that the researcher considers might help in giving an explanation of that price. Each characteristic or 'explanatory factor' has an associated numerical parameter, and the modelling works by varying the values of those parameters to get the best fit between the prices predicted from the explanatory factors and the actual prices. After experimenting with the form of the relationships between the price achieved at auction and the explanatory factors taken from the auction sales data, the final models I settled are ones from which statistically sound inferences can be drawn. Each artist was modelled separately, and the models takes largely the same form for all of them. Each is what is called a 'fixed effects' model in that the influence of each characteristic upon the price

achieved at auction is assumed to be fixed for the whole of the period of sales I have modelled. Although in practice it is likely that collectors' preferences have changed over time, the success of my models in giving predictions that are a good fit to actual auction prices indicates this is a good approximation for the collectors and artists I have modelled.

The interpretation of linear regression models needs to be handled with care. What is key, is that the model allows the researcher to give an answer to the questions posed. Faced with the same data researchers may develop different models suited to the questions they are asking and may find that different characteristics affect the price or that the same characteristic makes different contributions. Models need to be interpreted as a whole, with each characteristic considered in the context of the other characteristics included in it. The form of my models is one in which most of the parameters have an interpretation in terms of the percentage change in the price arising from a change in the associated explanatory factor, all other factors included in the model being equal.

The other key aspect to linear regression is the concept of statistical significance and the threshold to be used for it when interpreting the results of a particular model. To say that a potential explanatory factor to a model is not statistically significant is not to say that the estimated value of the associated parameter is very small, rather it is to say the uncertainty in the estimate is such that it cannot be confidently regarded as non-zero. For each parameter the model gives a 'p-value', or equivalently an estimated confidence level of one minus the p-value that it is statistically significant. In interpreting a regression model researchers may split explanatory factors into those whose contributions are and are not statistically significant at a certain threshold, using a confidence level threshold appropriate for the kind of work in which they are engaged. As I have discussed in Chapter 3, I have used a p-value of 0.1 in my work, or equivalently a confidence level of 90% that a parameter is statistically significant.

In this chapter my continued use of phrases such as 'in my model' and 'all other things being equal' may come across as repetitive and unnecessary but I have adopted this practice to highlight the model-dependent and holistic nature of the interpretation of regression models.

Modelling Results

A summary of the results from my modelling is given in Table 4.2. As can be seen, my models give an explanation of between 70% and 89% of the variation in the price achieved at auction for all artists, bar A. R. Penck.⁴⁹ In the cultural economics literature that I have surveyed and reviewed in the introduction to this chapter, regression models of the art market typically have an explanatory performance of only around 40%. The much better performance of my models is most likely attributable to my modelling of individual artists. David Galenson, who has modelled individual artists, Douglas Gordon, and Mathieu Aubry, Roman Kräussel, Gustavo Manso, and, Christophe Spaenjers, who have artist-specific explanatory factors in their models, have a similar level of explanatory performance to mine for their models.⁵⁰

Table 4.2 also identifies for each potential explanatory factor whether it is a statistically significant determinant of the price paid at auction for the artist being modelled and, if so, the percentage change in the price that can be associated with it.⁵¹ In it I have not followed the standard practice in the economics literature which is to give the values of the parameters in the model, but, for ease of understanding, I have presented the associated percentage changes in the price. The two are not identical and I explain why in Appendix B, where the parameter values are presented along with their confidence levels. I have also excluded the constant term in my models, although it was a significant factor for all artists. In some regression models the constant can be interpreted as a baseline price, however, in mine it does not admit of such an interpretation as its value is dependent upon other choices such as whether to measure size in square centimetres or square metres. It is for this reason I have not presented the constant in Table 4.2.

⁴⁹ As I discuss in Appendix B, the measure of explanatory performance, or ‘goodness-of-fit’, I have used is that of the correlation between the actual and predicted auction prices. A correlation of 100% happens when the predicted prices are identical to the actual. A correlation of 0% indicates that the model is of no value in predicting auction prices.

⁵⁰ Galenson, (1997), Gordon, (2011), and Aubry, Kräussel, Manso, and Spaenjers, (2019).

⁵¹ Factors which were not relevant to an artist and did not feature in the model for their auction prices are designated as n/a, or not applicable, in Table 4.2.

Difference in price associated with:	Alexander Calder	Sam Francis	Asger Jorn	Martin Kippenberger
Doubling of size of work of art	74.3% more ⁵²	46.2% more	70.7% more	94.7% more
Oil compared with other media	n/a	67.1% more	142.0% more	117.0% more
Sale at Sotheby's or Christie's compared with other auction houses	14.5% more	32.3% more	not sig	26.8% more
Sale in the United States compared with other country	17.9% more	not sig	n/a	61.0% more
Sale in the United Kingdom compared with other country	18.2% more	21.1% less	not sig	39.5% more
Sale in Hong Kong compared with other country	n/a	n/a	n/a	n/a
Sale in France compared with other country	n/a	n/a	n/a	n/a
Sale in native country compared with other country	n/a	n/a	not sig	not sig
Auction date, annual rate of appreciation	8.9%	1.7%	0.7%	8.0%
Specific title compared with generic title	11.1% more	37.3% more	22.4% more	21.5% more
Artist's age at execution of work	not sig	declines through to age 53 then picks up modestly	modest decline through to age 40 then picks up somewhat	declines through to age 36 then picks up
Explanatory performance	82%	78%	76%	75%

Table 4.2. Results from Regression Modelling.

⁵² For Alexander Calder, the increase in price reflects a doubling in length rather than a doubling in size.

Difference in price associated with:	Joan Mitchell	Yoshimoto Nara	Albert Oehlen	A. R. Penck
Doubling of size of work of art	68.6% more	64.2% more	53.4% more	46.8% more
Oil compared with other media	n/a	27.0% more	not sig	24.9% more
Sale at Sotheby's or Christie's compared with another auction house	not sig	not sig	21.4% more	not sig
Sale in the United States compared with other country	not sig	not sig	43.7% more	not sig
Sale in the United Kingdom compared with other country	not sig	not sig	45.6% more	not sig
Sale in Hong Kong compared with other country	n/a	not sig	n/a	n/a
Sale in France compared with other country	not sig	not sig	n/a	n/a
Sale in native country compared with other country	n/a	33.3% less	n/a	15.0% less
Auction date, annual rate of appreciation	10.9%	11.9%	15.2%	not sig
Specific title compared with generic title	not sig	60.4% more	25.7% more	14.6% more
Artist's age at execution of work	not sig	increases with artist's age through to 44 then declines	increases with artist's age through to 41 then declines	declines through to age 50 then picks up modestly
Explanatory performance	84%	79%	81%	60%

Table 4.2. Continued.

Difference in price associated with:	Sigmar Polke	Gerhard Richter	Cy Twombly	Christopher Wool
Doubling of size of work of art	83.3% more	82.5% more	79.4% more	71.3% more
Oil compared with other media	not sig	n/a	not sig	n/a
Sale at Sotheby's or Christie's compared with another auction house	not sig	23.7% more	not sig	27.5% more
Sale in the United States compared with other country	105.2% more	76.7% more	138.1% more	not sig
Sale in the United Kingdom compared with other country	127.2% more	69.2% more	115.3% more	n/a
Sale in Hong Kong compared with other country	n/a	n/a	n/a	n/a
Sale in France compared with other country	n/a	n/a	n/a	n/a
Sale in native country compared with other country	38.6% more	62.7% more	n/a	n/a
Auction date, annual rate of appreciation	9.4%	16.5%	10.7%	20.4%
Specific title compared with generic title	27.5% more	not sig	35.8% less	49.6% more
Artist's age at execution of work	declines through to age 59 then picks up somewhat	increases with artist's age through to 61 then declines	not sig	increases with artist's age through to 41 then declines
Explanatory performance	70%	89%	71%	71%

Table 4.2. Continued.

A comparison of Tables 4.1 and 4.2 shows that having a generic title and the 'other' factors are not included in the regression model. The reason is that when the sales data is divided into categories the model looks at the relative importance of those categories compared with a baseline that is excluded from it. So, for instance, the parameter associated with 'Sale at Sotheby's or Christie's' is interpreted in terms of whether sales at those auction houses achieve a statistically significant discount or premium compared with sales at other auction houses, all other factors being equal. The model does not include 'Sale at other auction house' as another explanatory factor. With the locations included in my models, the relevant comparison is with sales with all the countries not included in the model, which vary by artist. With Sam Francis, for instance, the other countries are Austria, Belgium, Denmark, France, Germany, Hong Kong, Italy, the Netherlands, Sweden, Switzerland, and the United Arab Emirates.

Size of Painting

As with many other studies, the models I have developed are ones where the size of a work of art is an important determinant of the price achieved at auction. It is a highly significant factor for all the artists I have looked at. In my models, for most artists doubling the size of their paintings increased the average price achieved at auction by over 70%, all other factors being equal. Collectors may consider that the size of a work is a sign of its quality, or may simply be prepared to pay more for a larger painting as it covers more wall space. As the sociologist Olav Veltius has discussed in his study of paintings sold by dealers in New York and the Netherlands, the supply side of the market may also be an influence, as it is typically costlier in both time and materials for an artist to produce a larger work than a smaller one.⁵³ These increased costs may be reflected in the prices charged to collectors when the work is first sold, and will subsequently influence the auction price when they are resold.

Comparing the details of my study with other investigations indicates that collectors' preferences have changed significantly. Size was valued much more by the collectors of the contemporary artists in the auction period I have examined than in the earlier contexts looked at in other studies. For instance in David Galenson's models of paintings sold at auctions of contemporary art from 1980 to 1997 by 42 predominantly American or American-based artists born before the Second World War, doubling the size of a painting leads to a price increase

⁵³ Olav Veltius, *Talking Prices: Symbolic Meanings of Prices on the Market for Contemporary Art*, (Princeton and Oxford: Princeton University Press, 2005), pp. 97 - 116.

of between 25% and 60% for the large majority of artists, all other things being equal.⁵⁴ The economists Richard Agnello and Renée Pierce have looked at 66 American artists born before World War Two and at auction sales over the period from 1971 to 1992. In their model, the impact of an increase in size upon the sales price is smaller than in mine.⁵⁵

Other studies have also found that collectors can be put off by very large paintings, and there is a maximum size beyond which the price paid at auction falls, all other things being equal. In Agnello and Pierce's study the maximum size is 6.53 square metres.⁵⁶ In their model of auction sales over the period from 1961 to 1990 of paintings by 82 French or Paris-based Impressionist, Modern and Contemporary artists and by 82 'Old Masters', the economists de la Barre, Duclou and Ginsburgh find that there is a maximum size of 5.89 square metres for the former and 1.70 square metres for the latter.⁵⁷ The usual explanation given in the economic literature for the presence of a maximum size is that not many private houses have the space to accommodate the largest works sold at auction.⁵⁸

To compare the preferences of collectors included in my modelling with those of earlier studies I looked at an amended model able to indicate a maximum size at which prices peaked for each artist if there were one. What this version suggests is that size was typically not a barrier to the collectors I have modelled. In those models, for only two artists was there a maximum size at which prices peaked within the range of sizes of their paintings sold at auction. For the other ten there was no maximum size, or the maximum size far exceeded that of any of their auctioned paintings. We might look to understand the differences between the results of my modelling and earlier studies as relating to change in the auction market. In recent years collectors have become more likely to put works of art purchased at auction into storage, especially if buying high value paintings for investment purposes.⁵⁹ The number of contemporary art museums has grown substantially in recent decades. In a 2015 global survey of private contemporary art museums, the market analysts Larry's List found that 53% had opened in the years from 2000 to 2010.⁶⁰ Both of these factors will have affected the level

⁵⁴ Galenson, (1997).

⁵⁵ Agnello and Pierce, (1996). Agnello and Pierce have one model which includes all 66 artists and so the size variable is an average across those artists. The form of their model is different to mine and the size-effect changes with the size of the painting. For a painting of average size, in their model a 1% increase in size results in a 0.31% increase in price. The comparable increases for my model are from 0.55% to 0.98%.

⁵⁶ Agnello and Pierce, (1996).

⁵⁷ de la Barre, Duclou and Ginsburgh, (1994).

⁵⁸ Agnello and Pierce, (1996), and Higgs and Forster, (2011).

⁵⁹ For a discussion of these changes in the market see Brook Mason, 'What is Driving the Soaring Demand for Art Storage?', *Apollo: The International Art Magazine*, 27 June 2017, pp. 67 - 68.

⁶⁰ Larry's List, *Private Art Museum Report*, (Hong Kong: Larry's List, 2015), p. 25.

of demand for the largest paintings sold at auction. We might also speculate that some collectors active at the top end of the contemporary art auction market over the last thirty years have also had more wall space to fill than those active in the earlier periods looked at in other studies.

Medium

The traditional hierarchy in the fine arts was that oil paintings on canvas were where a painter executed their most important and valuable works. This privileging of oil on canvas persisted within the dominant modernist aesthetic well into the twentieth century and so it is no surprise that in my model for two of the painters active during the middle decades of the twentieth century, Sam Francis and Asger Jorn, collectors have valued their paintings in oil much more highly than those in other media, all other things being equal. For Francis the premium is 67.1% and for Jorn it is 142.0%. In their model Agnello and Pierce also find that a substantial sized premium of 75.1% was given to paintings executed in oil across the artists and auction sales included in their analysis.⁶¹

From the 1950s onwards these aesthetic distinctions were increasingly challenged as more artists worked across multiple media and genres, adopted emerging technologies such as film or video, or claimed that what was of artistic value was the idea rather than the execution. Since then, a common critical trope has been to proclaim the 'death' or the 'resurrection' of painting. Despite these fundamental aesthetic changes, some collectors of contemporary artists active in recent decades appear to have persisted in preferring paintings in oil above paintings in other media. In addition to Francis and Jorn, for three other artists their paintings in oil have sold for significantly more than works in other media. For instance, the 12% of sales identified in my auction data as being of works executed in oil by Yoshimoto Nara were for prices 27% more on average than the sales prices for his paintings in other media. Ironically, for Martin Kippenberger, an artist who challenged and satirised the art historical tradition, in my model collectors have paid more than double for his paintings executed in oil on average than for those in other media.

⁶¹ Agnello and Pierce, (1996). Van Rooneborg and Spanjaers, (2013), also find that works of art executed in oil command a premium. However their base of works includes paintings, prints and drawings so it is not possible to draw any conclusions on paintings in oil compared to paintings in other media. The other studies I have reviewed also do not look at paintings executed in oil compared with other media.

Auction House and Location of Sale

My models, as other studies, indicate that both the location of the sale and the auction house can make a substantial difference to sales prices all other factors being equal. For seven artists, consigning a work for sale by Christie's or Sotheby's delivered a premium to sellers compared with other auction houses. With works by six artists, a painting consigned in the United States sold for more, on average, than a painting sold in one of the locations not included in the model. For largely the same group of artists, selling in the United Kingdom delivered a similarly-sized premium, although with San Francisco, paintings sold for less in the United Kingdom than the other locations included in my model.

Collectors may have considered that being sold by Sotheby's or Christie's was a signal of a work of art of the highest quality. We can also understand the premiums associated to auction house and location as relating to the structure of the high end of the contemporary art market over the years I have looked at. For all the artists I have modelled bar Yashimoto Nara and Asger Jorn, Sotheby's and Christie's were, and remain, their most important auction marketplaces. Both auction houses have a global presence with offices in New York, London, Paris and Hong Kong along with other locations. Their sales of contemporary art are heavily promoted, and the auction houses may also deal directly with potential buyers of high-priced works. Buyers from around the world or their representatives bid at these auctions in person, on the phone or online. For most of the period of auction sales I have modelled, the most important events have been the contemporary art weeks held in New York and London twice a year, when both Sotheby's and Christie's will hold day and evening auctions. In recent years Phillips has also looked to compete directly with Sotheby's and Christie's holding its own events during contemporary art week, and smaller auction houses will also arrange their New York or London sales of contemporary art to tap into this demand. In my model we can see that the marketing and promotional efforts of the auction houses have paid off. Sales at Christie's and Sotheby's often resulted in higher average sales prices compared with other auction houses, as did auction sales in the United States and the United Kingdom compared with those in other locations. It is also notable that not only have largely the same artists achieved higher sales prices in the United States and the United Kingdom, but the premiums attached to those two locations were very similar, which suggests that for those artists the levels of demand in the two markets have been commensurate.

My model results for Yoshimoto Nara give a window into how the auction market for contemporary art has been changing. Over the last fifteen years sales in Asia, in particular of Asian artists, have grown substantially and the auction market for contemporary art in China, including Hong Kong, is on one measure now bigger than that in London.⁶² Since 2017 there has been a contemporary art week held twice a year in Hong Kong. Sotheby's and Christie's have a smaller market share in Asia than in their traditional markets, where several Chinese auction houses have comparable levels of sales. Yoshimoto Nara is one of thirteen Asian artists, mainly Chinese, who were in the top 50 best-selling post-war contemporary artists at auction in 2018/2019.⁶³ In my models, sales of paintings by Nara at Sotheby's or Christie's do not command a premium compared with other auction houses, neither do sales in the United States or the United Kingdom compared with sales in Hong Kong and in mainland China.

In the economist Rustam Volikov's regression model for auction sales over the period from 1985 to 2013 of sculptures by 181 artists active from the early nineteenth-century to the present day there is a clear positive home bias.⁶⁴ Sculptures sold for more on average in the artist's native country than elsewhere, all other things being equal. For artists with an international reputation Volikov attributes the domestic premium primarily to the patriotism of some collectors. Volikov's results cannot be directly compared to mine as he looks at an average across all the artists whose sales he has modelled and at a different group of artists and period of sales. However, my models do suggest that when looking at artists individually the results can be more mixed. There was no clear pattern of home bias one way or the other amongst the collectors of the six artists where I have been able to look at sales in their native country. For Sigmar Polke and Gerhard Richter, sales in Germany boosted average prices compared with the locations not included in my models, but with the former to levels below those for sales in the United States and the United Kingdom. For Yoshimoto Nara and A. R. Penck, sales in their home countries were for lower prices on average than those in the locations not included in my models. Sales of paintings by Martin Kippenberger and Asger Jorn in their home country did not have a significant impact upon the price achieved at auction.

⁶² See <https://www.Artprice.com/Artprice-reports/the-contemporary-art-market-report-2019>. Artprice defines contemporary art as that produced by artists born after World War Two and so will exclude works by artists such as Gerhard Richter and Cy Twombly whose auction sales remain concentrated in the United States and Europe. If these artists are included, sales of contemporary art in London continue to exceed those in China.

⁶³ See <https://www.Artprice.com/Artprice-reports/the-contemporary-art-market-report-2019>.

⁶⁴ Rustam Volikov, (2015). Volikov looks at auction sales from 1985 to 2013 covering 181 artists from 28 countries active from 1800 onwards. In contrast to my model, Volikov only compares the native country of each artist with all other countries, whereas I have included sales in several countries separately in my model.

Rates of Appreciation

The auction market has its ups and downs, and there can be significant fluctuations in the prices achieved at auction for particular artists as they come into and out of fashion. As I have already discussed, and cover in more detail in Appendix B, I made adjustments in my models for these factors and so my results on the rate of appreciation can be understood as representing the underlying trend in the real US Dollar value at auction of works by each artist. In my models, the prices achieved at auction by ten of the twelve artists have appreciated considerably over the period I have looked at, increasing by 8% or more per annum on average in real US Dollar terms. Accumulated over the period of sales I have modelled these increases are substantial. With Gerhard Richter, for instance, real prices for his paintings increased nearly thirty-fold from 1987 to 2019.

As with my reading of the results on auction house and location, these results suggest further insights into the auction market for contemporary art. All these artists have been in high demand and have been consistently amongst the best-selling artists at auction. All are examples of the 'virtuous' circle of auction sales, high-end dealers, art museums and major art fairs, biennales and other exhibitions which have driven prices in the contemporary art world. The other three artists, Sam Francis, Asger Jorn, and A. R. Penck, where the rates of appreciation are substantially lower have a much smaller art world 'footprint' than the other nine. Compared with these other artists their paintings are more likely to appear for sale during an auction house's daytime sales rather than the more prestigious evening sale. They are less likely to be represented by one of the leading international art dealers. Compared with the other ten artists bar Yashimoto Nara and Albert Oehlen, their art institutional presence in major collections or through retrospectives at major art museums is also lower. A. R. Penck is an artist who has come into and out of fashion more than any of the other artists I have modelled. Prices for his paintings at auction have been volatile and, in my models, there has been no underlying change in the real value of his paintings all other factors being equal.

Age of Artist

Before discussing the market value collectors have ascribed to different types of title, I will look at the impact of the age of the artist at execution on the price paid at auction for a work of art. Of all the characteristics of paintings I have modelled, this is probably the most difficult to interpret and explain. The form of the model I have used means the age profile should be seen as showing the broad trend in prices paid at auction for works and should not be interpreted on a year-by-year basis. As will be seen, there are some suggestive patterns in

the age profiles across artists. However, although these merit further investigation, that is beyond the scope of this thesis and in what follows I will only offer some speculative comments, in particular where my results relate to those of other scholars. To develop a full explanation would require a detailed study of the biographies of each artist and of the ways in which their artistic careers have been constructed and presented by art historians, critics, curators and the main auction houses. The inter-relationship between the age of the artist and the historical periods during which they worked would also need to be considered. Some collectors may, for instance, have a preference for works from a movement associated with a particular period rather than for works executed at a particular life stage of an artist involved in that movement. A further complicating factors is that most of the artists were alive for some or all the period of sales I have modelled, and so works executed at different ages would have come onto the market at different times during that period. A painting executed early in an artist's career and brought to market at that time might command a lower price than the same work or a later painting sold once they had an established reputation. Looking at a larger number of artists would also help in confirming the strength of the patterns which can be seen with the twelve I have modelled.

The age profiles in my models can be spilt into three groups depending upon the life stage of the artist. Chart 4.1 gives the age profiles for Alexander Calder and Asger Jorn, who both died in the decade prior to the start of the period of auction sales I have modelled. As can be seen, collectors were indifferent between Calder's mobiles executed at different points during his career. For Asger Jorn, collectors had a modest preference for paintings executed towards the end of his career.

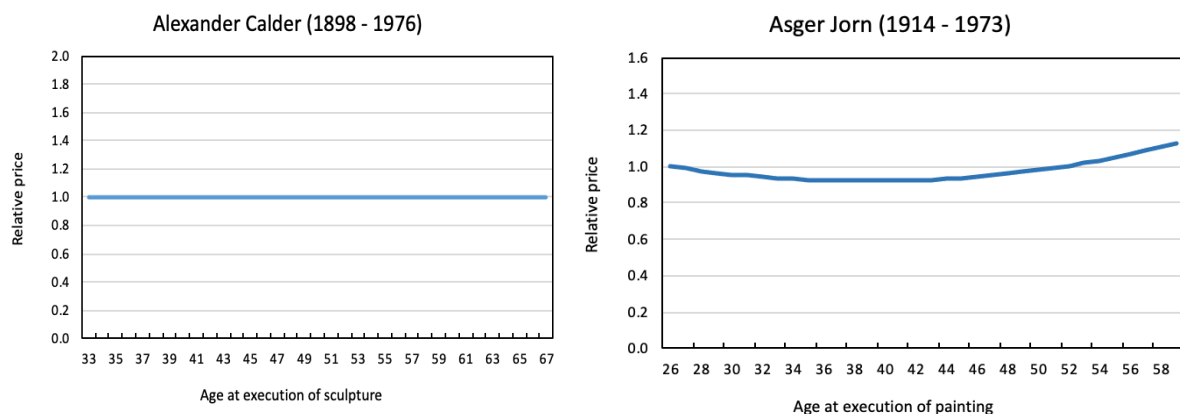


Chart 4.1. Relative prices for paintings executed at different ages by Alexander Calder and Asger Jorn.

Six of the artists I have modelled died during the period of auction sales included in my analysis. Their age profiles are given in Chart 4.2, which shows that collectors of Joan Mitchell and Cy Twombly were indifferent between paintings executed at different stages of their careers. For the other four artists, collectors had a preference for early-career paintings compared with mid-career works. In my models, mid-career works by these artists sold, on average, for around one-third to one-quarter of early-career works, all other things being equal. For all these artists prices changed little for mid-career works before picking up modestly for paintings executed at the end of their careers.

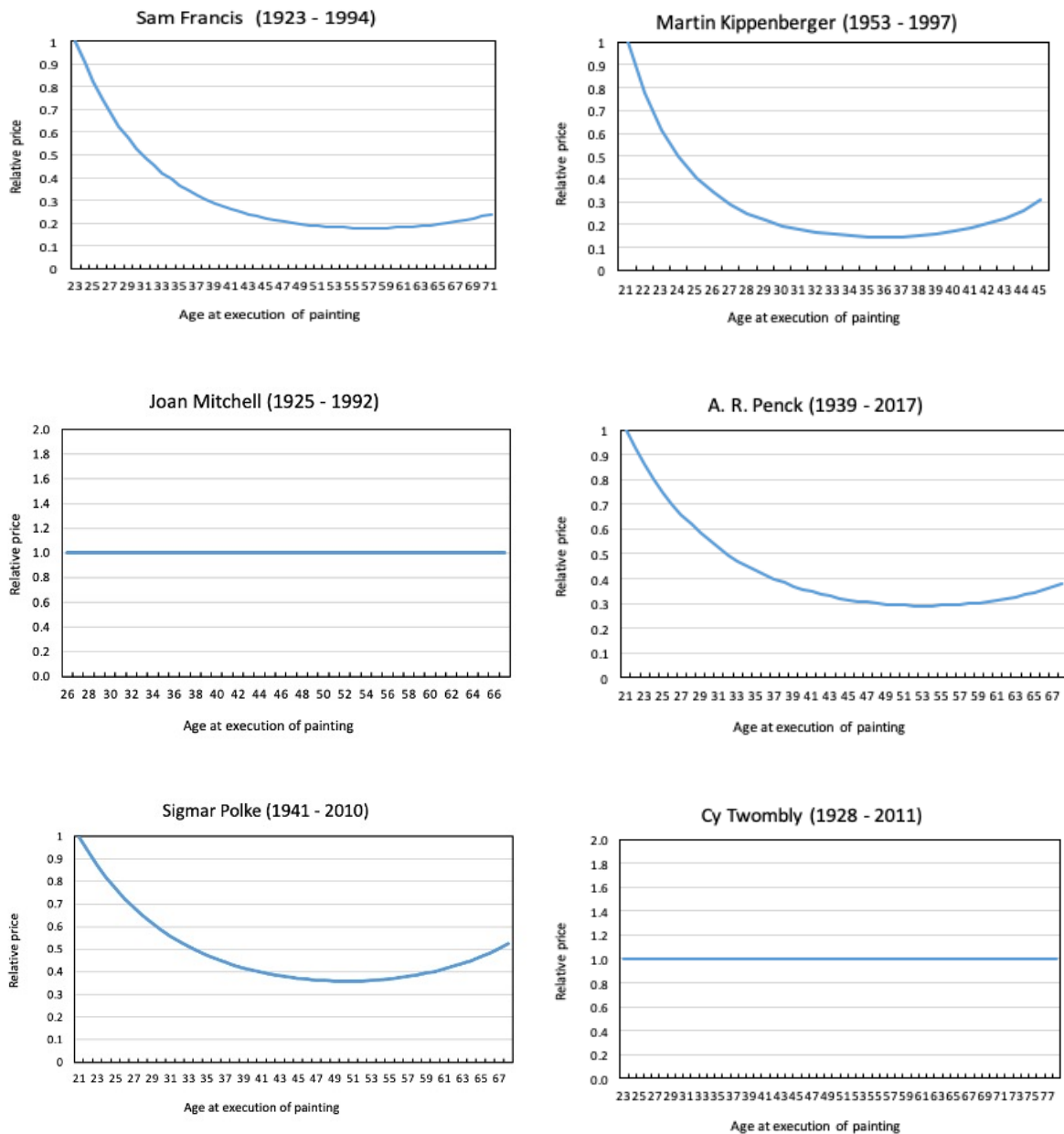


Chart 4.2. Relative prices for paintings executed at different ages by Sam Francis, Martin Kippenberger, Joan Mitchell, A.R. Penck, Sigmar Polke, and Cy Twombly.

David Galenson's modelling highlights the need to consider the period when a work was executed alongside the age of the artist when looking to understand prices paid at auction. In his models of 42 American or American-based artists, the peak ages for the older artists were higher than those for the younger artists. In both cases, these were during the years from the late 1940s to the early 1970s, a period when American art has been widely seen as innovative and world-leading.⁶⁵ The highest prices paid for paintings by Sam Francis and Sigmar Polke were for paintings executed in the 1950s or 1960s, and suggest that some of the more recent cohort of collectors I have modelled may have continued to value art from the main movements of that time. Sam Francis is often associated with Abstract Expressionism or Tachisme, and Sigmar Polke with a European version of Pop Art.

It is widely believed auction prices increase once an artist dies, a phenomenon that has been confirmed in the work of several cultural economists.⁶⁶ This means that paintings executed late in an artist's career may first come onto the auction market once collectors have the expectation they may soon increase in price. These expectations would boost the average price of later works compared with earlier ones, all other things being equal. Francis, Penck and Polke all had lengthy careers and died during the years covered by my auction sales data and so the pick-up in the average price for their late career works may reflect that factor. Martin Kippenberger died at the premature age of 44 and so a different explanation of his age profile is called for. Paintings executed in the last few years of his life appeared on the market in the late 2000s by which time his critical and market reputations were established, and so would have been boosted in price compared with mid-career works sold in earlier years. The age profile for Joan Mitchell does not show an upturn for paintings executed towards the end of her career. One possible reason is that Mitchell died in 1992, and, as I have already discussed, the large majority of the sales of her paintings have been in the current century.

⁶⁵ Galenson, (1997).

⁶⁶ See, for instance, Heinrich W. Ursprung and Christian Wiermann, (2008).

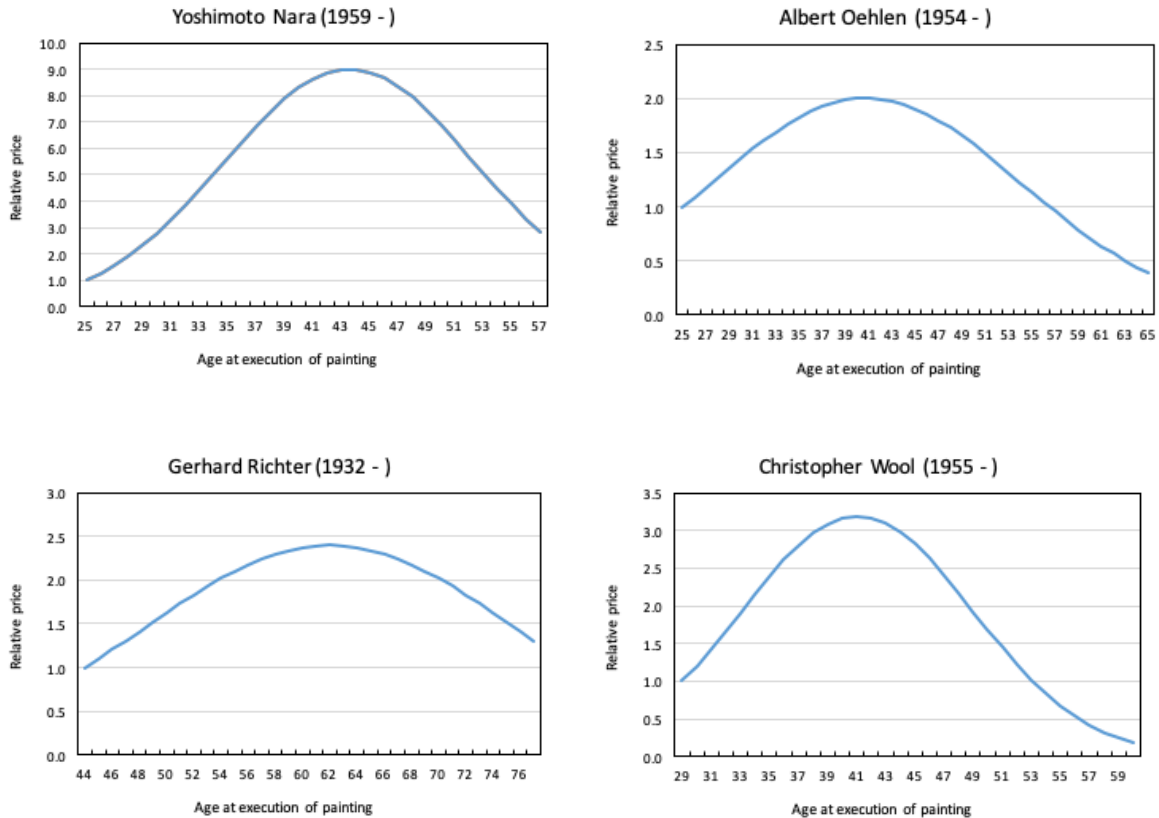


Chart 4.3. Relative prices for paintings executed at different ages by Yoshimoto Nara, Albert Oehlen, Gerhard Richter, and Christopher Wool.

The remaining four artists - Yashimoto Nara, Albert Oehlen, Gerhard Richter and Christopher Wool - were all living at the end of the period of sales I have modelled. Their age profiles are given in Chart 4.3, and as can be seen the profiles for these artists have similar shapes. For all four artists, early works have sold on average for less than mid-career paintings, all other factors being equal, and there is a peak age for each artist beyond which prices go into decline. For Nara, Oehlen and Wool, who are among the youngest artists in my analysis, prices peak for paintings executed in their mid-40s.

The price profiles for the artists who were alive at the end of the sales period I have modelled are very different from those of the artists who died before or during it. They also differ from the age-profiles for artists in the historical contexts examined by David Galenson and Douglas Gordon.⁶⁷ In Galenson's study there is no correlation between age-profiles and the life stages of the artists in the period of auction sales he models. In Gordon's study of Canadian artists the age profiles all have an inverted-u shape, as with the age profiles shown in Chart 3.2, and the age at which prices peak tends to decrease the more recently the artist was born.

⁶⁷ Galenson, (1997), Gordon, (2011).

However, developing an understanding of why these profiles have the shape they do, and why they differ from the profiles given in earlier studies is, as I have already mentioned, beyond the scope of this thesis.

Type of Title

In their studies Renneboog and Spaenjers, and Aubry, Kräussel, Manso, and Spaenjers developed regression models in which works of art that are untitled or had generic words such as 'abstract', 'portrait' or 'landscape' in their titles sold for less at auction, all other things being equal, than works without such words in their titles.⁶⁸ However, the base of works of art in both studies is so broad, including paintings, drawings, prints, and editions, and the number of artists so large at over 100,000 that it is not possible to draw any conclusions regarding collectors' preferences with paintings or sculptures alone or with artists active in a particular period or movement.

My models, in contrast, are more focused and in them there is a clear pattern of collectors having preferred paintings or sculptures with one type of title. It largely confirms the working hypothesis I set out in the introduction, that collectors tend to have a preference for works of art with specific titles. For nine of the twelve artists, collectors paid more on average for works presented at auction with specific titles than for comparable works presented with generic titles. The premiums paid for works with specific titles varied from 11.1% with Alexander Calder's mobiles to 60.4% with paintings by Yoshimoto Nara. However, you also need to look at the reasons why collectors of the three other artists I have modelled did not exhibit such a preference.

For Gerhard Richter's abstracts, collectors have been indifferent to the type of title it has, paying the same on average, all other factors being equal, for paintings titled 'Abstraktes Bild' and those with a specific title. In contrast to the other artists I have modelled, where the literature is heavily focused on works with specific titles, Richter's use of 'Abstraktes Bild' as a title is a key feature of some readings and critics are divided in their opinions of his best abstracts between ones with specific titles and ones with generic titles. The art historian Christine Mehring, for instance, rejects the standard English translation of 'bild' as 'painting' as misguided and leading to misinterpretation, commenting that the German term is complicated and that the basic sense it has as 'picture' or 'form of representation' is what Richter seems to have in mind in his remarks in this context.⁶⁹ The idea of Richter's abstract

⁶⁸ Renneboog and Spaenjers, (2013). Aubry, Kräussel, Manso, and Spaenjers, (March, 2019).

⁶⁹ Christine Mehring, 'Richter's *Willkür*', *Art Journal*, Winter, (2012), pp. 21 - 22.

works as pictures is central to her interpretation. Mehring also identifies three paintings, all titled *Abstraktes Bild*, as being among his best and most complex abstracts.⁷⁰ Individual titles play a more important role in the curator Mark Godfrey's interpretation of Richter's abstracts. In his presentation to a 2012 Tate Modern symposium on Richter which accompanied the *Panorama* retrospective, they are seen as singling works out from his ongoing series of abstracts as being his most important works and worthy of special attention.⁷¹ As with collectors who value specific titles, critical opinions may have influenced collectors of Richter's abstracts to some extent but it is unlikely to be the complete explanation.

I would speculate that the generic title 'Abstraktes Bild' has also taken on a 'brand' value in relation to Richter's abstracts which simply presenting a work as 'untitled', which is predominantly the case with the generic titles of the other artists I have modelled, does not allow. The term 'Abstraktes Bild' is often used not only as the title of a single work by critics but, in the plural, as a name for all his abstract paintings however they are titled. In my model, collectors have been indifferent between a 'branded' Richter abstract and one with a specific title.

Collectors have also shown no preference between generic and specific titles for Joan Mitchell's paintings. With Cy Twombly, the final artist I have modelled, paintings with specific titles sold for much less than those with generic titles, all other things being equal. None of the explanations I have just offered can account for these preferences. The critical literature is biased towards Mitchell's and Twombly's paintings with specific titles. Mitchell's and Twombly's paintings with generic titles are predominantly presented at auction as untitled. The paintings of both artists are in high demand but, if the underlying rates of appreciation in their auction prices is taken as a measure of that demand, no more so than several of the other artists I have modelled. The supply of paintings by both artists to the auction market as measured by the number of sales is also at a similar level to that of other artists. I remain unable to even venture an explanation of the preferences collectors have had for Mitchell's and Twombly's paintings with different kinds of title.

⁷⁰ Mehring, (2012), p. 23.

⁷¹ Mark Godfrey, Tate Modern Panorama Symposium, April 4, 2012.

My investigation also shows the importance of using regression models to give an understanding of collectors' preferences. With Martin Kippenberger, paintings with generic titles have sold for nearly three times more, on average, than those with specific titles. However, collectors have had a modest preference, all other things being equal, for the latter. Of two paintings similar in all respects except that one had a generic title and the other a specific title and on sale at the same auction, the latter would sell for more than the former. In my model there is no one factor that largely accounts for this difference. Rather, Kippenberger paintings presented at auction with generic titles are larger than those with specific titles, are more likely to have been sold at Sotheby's or Christie's and in the United States, and are more likely to have been executed in oil. All these factors boost the average price of his paintings in my model compared with paintings having specific titles.

For Gerhard Richter, as shown in Table 4.2, paintings with specific titles have sold for just over twice the price, on average, of those with generic titles. However, in my model collectors were indifferent between the type of title of a Richter abstract has, all other things being equal. What accounts for the divergence in average price is mainly the difference in size between paintings with specific titles and those with generic titles. Richter paintings with specific titles are 2.2 times the size of those with generic titles on average and in my model this is associated with a doubling in the price.

4.4 Summary and Conclusions

Overview

In this chapter I have set out the second of my case studies of the use of quantitative techniques and the utilisation of digital resources in the history of art. It builds on my first case study in exploring different resources and techniques, and in examining a different historical and institutional context. In it I use the explanatory statistical technique of regression analysis to model auction sales for twelve contemporary artists. These models give answers to the questions of which characteristics were important to collectors and of how strongly they valued them. They also give some insights into the operation of the auction market.

In summary, size was an important driver of the price paid at auction for all of the artists I have modelled. Paintings in oil have continued to appeal to some collectors and even works in oil by artists who consciously challenged art historical traditions such as Martin Kippenberger commanded a premium at auction. Although the number of artists I have looked at is small, there are some strongly suggestive patterns in how the age of the artist at execution affected

the price achieved at auction, which might merit further investigation. For nine of the twelve artists there was a clear preference among collectors for paintings presented at auction with specific titles. I speculatively attribute these preferences to the greater 'seductive' power of specific titles compared with generic titles.

My models indicate that consigning a painting for sale by Sotheby's and Christie's delivered a premium to sellers, as did selling in what were the two major centres for the contemporary art auction market, New York and London. For some artists, sales in their home country delivered premium, but for others it did not. We can also see how the 'virtuous circle' of the major auction houses, art fairs, dealers and art museums in the contemporary art world has driven prices in the auction market.

Digital Art History

The work I have presented in this case study adds to the body of scholarship in digital art history which has looked at auction sales or has used regression modelling. It is of art historical value to have an understanding of what motivates collectors of art and of their preferences between paintings, as it is of how the auction market functions in the increasingly globalised art world of the last few decades. Regression modelling allows for measurable or classifiable factors that might influence collectors or the market more generally to be investigated, all at once and taking account of their inter-relationships.

The approach I have followed is one which is well-established with cultural economists as a way of modelling and understanding the art auction market. I have extended it to include the type of title a work was presented with at auction as one characteristics, in addition to those which are standardly included in such models. My work in looking at the preferences of collectors of individual artists for works with different kinds of title is new in cultural economics. My readings also complement and re-contextualise earlier studies, showing how collectors' preferences and market structures have changed. For instance, size was a stronger driver of the auction price in my work and, in contrast to earlier studies, collectors were not put off by very large paintings. The 'death effect', where cultural economists have confirmed the belief that prices rise after the death of an artist, may be seen in my models in an upturn in prices for works executed late in the artist's career.

The method of regression analysis has allowed me to bring a new kind of knowledge into art history. Neither traditional art historical methods nor looking at average prices for each factor on their own allow the art historian to develop the kind of disaggregated understanding of collectors' preferences I present in this chapter. It also shows how a way of thinking that has not conventionally been used in the discipline, namely 'all things being equal', or 'ceteris paribus', where changes in one factor are considered in isolation from changes in the others, can contribute to art history enquiry.

In addition, the extensive body of work by cultural economists I have drawn on for this case study appears to have had very limited exposure within art history. I would hope that this thesis will bring that work to the attention of art historians, alongside my own contribution.

5. MODERN AND CONTEMPORARY ART

5.1 Introduction

In their studies into the history of titling the art historians Stephen Bann and Ernst Gombrich read their way chronologically through a selection of artists providing their interpretations of how titles contributed to the meaning of the works they named.¹ Both accounts focus heavily on artists from the history of Modernism. Both argue convincingly that, as Bann characterises it, 'a lot was at stake with titling', a site where the 'theoretical self-consciousness of the modernist painter expressed itself ... in an awareness of the utility of names' and in a 'diversity of titling practices'.² Bann begins his account of Modernist titling by looking at the Cubists. George Braque and Pablo Picasso 'palliated the ambiguity' of paintings which provided few of the 'conventional visual clues' to the viewer through the use of titles that followed a 'strict generic consistency'.³ However it appeared to the viewer, a Cubist painting was a 'Head of a Woman' or a 'Still-life'. For Gombrich, Pablo Picasso used titles such as *The Accordionist* to make the search for a fit between word and image an integral part of the viewing experience. Bann observes that Marcel Duchamp ran 'the gamut of challenging and transgressive uses of the title' in which they entered an 'unstable zone of meaning'.⁴ On Gombrich's reading, Joan Miro's titles such as *Le Crépuscule rose caresse le sexe des femmes et des oiseaux* [The Pink Twilight Caresses The Sex of Women and Birds] 'suggest the workings of the Freudian primary process'.⁵ Bann's brief account of Modernist titling ends with Paul Klee, and he reads the artist's practice of inscribing a painting's title on the canvas as playing upon 'the containment of the Iconic within the Symbolic as a central component of its desired interpretation'.⁶ Gombrich's finishes his survey by looking at several Abstract Expressionists. Mark Rothko, for instance, with titles such as *Dark on Brown, No. 14*, is read by Gombrich as intending to 'induce the contemplative mood'.⁷

Bann observes that looking at titles is one way of 'retracing the history of modern art as a whole'.⁸ And in his book-length study of titles the art historian John Welchman develops that

¹ Stephen Bann, (1985), and Ernst Gombrich, (1985).

² Bann, (1985), p. 182.

³ Bann, (1985), p. 182.

⁴ Bann, (1985), p. 183.

⁵ Gombrich, (1985), p. 239. In his paper, Gombrich gives titles in their original language. I have followed his usage.

⁶ Bann, (1985), p.184.

⁷ Gombrich, (1985), pp. 240 - 241.

⁸ Bann, (1985), p. 184.

position in much more detail than the earlier two authors. His overarching aim is to 're-read some of the history and theory of Modernism' through the title as a 'site for avant-garde innovation and challenge to traditional theory and practice'.⁹ For Welchman the modern title has had 'a thousand faces' and he looks at over two hundred artists, grouped by the artistic movements and tendencies with which they were associated from Impressionism in the 1870s to Appropriation Art in the 1980s. He places the birth of the modern title with Symbolism and Post-Impressionism as artists consciously engaged with the opportunities titling offered. James McNeill Whistler was one of the first artists to use colour words in titles to indicate that the formal elements of the painting should be prioritised. Artists in Dada and Surrealism 'radically opened up and fervently disputed ... the signifying space of the title'.¹⁰ Artists involved with Conceptualism and with Institutional Critique interrogated and foregrounded the function of the title itself, instigating the meta-title.

Cutting across his chronological account, Welchman traces two themes in the use of titles. He looks at how titles were used in relation to ways the idea of 'composition' was adopted and contested by artists. Its importance was signalled in being used in titles by several artists associated with abstraction including Wassily Kandinsky and Piet Mondrian. Later artists used titles as part of their challenge to the 'compositional order' established by those earlier artists, as with Robert Rauschenberg's concept of the 'combine' - a word he used in the titles for a series of works from the 1950s. The second theme is 'minimal' titling in which artists gave non-referential titles to their works such as numbers or dates, or presented their works as untitled. Welchman looks at several artists who used numbers in their titles including Clifford Still, whose stated intention in using numbers as titles was to allow the painting to speak for itself. As one example of an artist who presented work as untitled, Welchman argues that Cindy Sherman's 'titular restraint' with her untitled film stills, along with her refusals to be interviewed or write about her art, opened those works up to multiple interpretations.

In this chapter I set out a different kind of account of what titling can tell us about the history of modern and contemporary art. Rather than looking at individual artists or movements and giving close readings of the ways titles contributed to the meanings of the works they named, my account adopts a synoptic viewpoint. As self-contained textual units, titles are well-suited to the sorts of sophisticated statistical approaches to text mining used by literary historians such as those I introduced in the literature review. Looking at titles in aggregate allows me to bring these predictive statistical techniques into art history and to explore the ways titles have

⁹ Welchman, (1997), p. 1.

¹⁰ Welchman, (1997), p. 209.

been used at a large scale. My approach is also to look, as far as possible within the constraints given by data availability, at a wide scope both geographically and temporally and in a way that does not privilege any of the countries or artistic centres represented in my dataset. My readings avoid what the art historians Catherine Dossin and Béatrice Joyeux-Prunel have characterised as the 'hierarchisation and exclusion upon which the modernist story is traditionally built'.¹¹ Bann, Gombrich and Welchman, for instance, all focus heavily on European or American artists with those from Latin America, Asia and other countries largely excluded. Such largely unexamined assumptions have persisted in many accounts of modern and contemporary art.¹²

To develop the readings I offer in this chapter I assembled a dataset of the titles and other metadata for over 59,000 modern and contemporary works of art created in the years from 1900 to 2009, and downloaded from the online collections of 35 of the world's major art museums. In the next section I give a critical review of these data sources and of the approach I took to building the dataset. As will be seen, this data is inherently 'noisy' and biased in that, for instance, the titles recorded may not be the original titles of the works in question, some artists are very heavily represented in some collections, and titles may include curatorial additions to the original. Drawing together data from multiple sources will have introduced some inconsistencies in my dataset. My transformations of the titles added further noise, for instance I translated all non-English titles into English. As I will come on to show, my statistical techniques and interpretive approach provide ways of dealing with some of the sources of noise and bias. In addition, the readings I develop are consistent with the canonical history of modern and contemporary art. Both give comfort that those readings are sound.

I then bring a number of statistical techniques individually and in combination to bear on this data to look at the question of what looking at titles in aggregate can tell us about the history of modern and contemporary art. My approach is to utilise those techniques to identify long-run trends and other patterns in the language used in titles, and to develop art historical

¹¹ Catherine Dossin and Béatrice Joyeux-Prunel, 'The German Century: How a Geopolitical Approach Could Transform the History of Modernism', in Thomas Dacosta Kaufmann, Catherine Dossin and Béatrice Joyeux-Prunel (eds.), *Circulations in the Global History of Art*, (London and New York: Routledge, 2015), pp. 183 - 201.

¹² Until recent editions, *Art Since 1900* looked predominantly at developments in Europe and the United States of America and provided an inadequate treatment of modern and contemporary art in most of the remainder of the world. Yves-Alain Bois, Benjamin Buchloh, Hal Foster, and Rosalind Krauss, R., *Art Since 1900: Modernism, Anti-Modernism, Postmodernism*, (New York and London: Thames & Hudson, 2004). David Hopkins frames his geographical account almost entirely in terms of the relationship between Europe and the United States of America. David Hopkins, *After Modern Art: 1945 - 2017*, (Oxford: Oxford University Press, 2018).

interpretations of them. As will be seen, the techniques I use complement each other and support a reading that provides a fresh perspective to that given in conventional art historical scholarship. I read the language used in titles in terms of a large-scale narrative as artistic interests signalled through the language used in titles came and went and were re-inflected, and as epistemic perspectives on the kinds of knowledge art can or should engender changed.

Coupled with my dataset, my analysis and reading of the language used in titles provides an interpretive framework that can be put to uses other than looking at titles in aggregate. I use it to consider the question of what text mining metadata can say about the circulation of ideas within modern and contemporary art. Splitting my data by country or region, I look at whether the interests signalled through the language used in titles were more important, emerged earlier or persisted longer with artists from some countries rather than others.

I then apply my interpretive framework to titles split between male and female artists, basing my assignment of gender on that given in the online collection or on the artist's given name. As I discuss in the Introduction our contemporary understanding is that gender is more complex than the simple binary of male and female. Indeed, contemporary artists have increasingly engaged with issues of gender, using their practice to express their gender identity, to question gender norms, and to challenge traditional representations of gender. Art historians have written about these artists or have taken queer or transgender theoretical perspectives on the history of art.¹³ However, the binary of male-female remains a useful category for historical analysis, for instance when looking at large-scale trends in contexts where it was the dominant understanding of gender, as is the case with this case study. Although this section is short, it provides a different perspective from which to compare some of the ways male and female artists have engaged in modern and contemporary art to that given by other scholars.

In the concluding section I draw together and summarise the readings I have set out and reflect on the art historical conclusions and methodological implications of my work. One important question is that of how the trends I identify and interpret relate to the earlier literature on titles and to the canonical history of modern and contemporary art. I will also address this question in the concluding section of this chapter.

¹³ For a survey see Catherine Lord and Richard Meyer, *Queer Art and Culture*, (2nd ed.), (London: Phaidon Press, 2019), and for an application of transgender theory see David J. Getsy, *Abstract Bodies: Sixties Sculpture in the Expanded Field of Gender*, (New Haven and London: Yale University Press, 2015).

The metadata provided in online museum collections is an underused resource in digital art history. From the methodological perspective, my work in this case study is innovative in that it uses a range of techniques in combination to develop a much richer interpretation of that data than those provided by other scholars working in digital art history. Looking at titles in aggregate and in quantitative terms has enabled me to address questions that cannot be answered with qualitative approaches to art history. There are also lessons for scholars working in the digital humanities from my approach in this case study.

The techniques I use also present substantial methodological and interpretive challenges. A good understanding of these techniques allows the art historian to get the best out of them and to avoid misinterpreting the results. As in the other chapters of my thesis, I do not take my methods as givens but the main text I engage in a critical review of the underlying assumptions and interpretive issues involved in their use. In Appendix C I provide more detail on my methods and on the steps I went through in developing the statistical models I present in this chapter. My aims are to give the interested reader an overview of the technical aspects of these techniques and examples of model development that may be of value for their own work.

5.2 Data Sources and Dataset Construction

Art museums are increasingly making their collections available online, and, although some major national museums had yet to do so when I was compiling my dataset, including the National Galleries of Modern Art in Italy and India, those that do allowed me to put together a dataset providing a broad geographical spread. For my dataset I collected metadata from the online collections of 35 art museums in 20 countries covering Europe, Asia, North America, and Latin America.

In developing the dataset I faced a number of definitional questions and issues around selection and categorisation. The first was what to classify as modern or contemporary art. As with my other case studies, my solution was institutional. I took as modern or contemporary art that displayed by art museums naming themselves that way or as defined through institutional or curatorial policies and the accumulated body of work in their collections labelled as such. The collection of The Metropolitan Museum of Art in New York, for instance, includes objects from prehistory to the present, and the data I gathered was on modern and contemporary art as defined by the institution in their online collection. The 35 institutions whose collections are included in my dataset represent all of those I was able to locate.

The modern and contemporary works included in most of the collections I sourced for my dataset are predominantly from the twentieth and twenty-first century and so for my analysis I restricted my dataset to work of arts created in the years from 1900. As you would expect, the number of works created in the 2010s in those museum collections was much smaller than for other periods, and so I excluded works of art created in that decade.

I also had to decide which types of work to include in the dataset. Where separated out by the institution in their online collections, as was the case with almost all the art museums I examined, I collected data on works classified as paintings, sculptures, installations or works in new media. The reasons for this choice are that these kinds of work typically have their own titles, whereas that was not always the case with other types of work such as photographs or prints. I could have limited my dataset to works classified as paintings and sculptures on the grounds they were the only types of work produced throughout the period I have considered. However, artists have increasingly worked across multiple media, and so this approach would have reduced the number of works in recent decades. It would also have introduced inconsistencies into my dataset as curatorial classifications can also vary. For instance, similar works by the same artist in my dataset have been categorised in one collection as a sculpture and in another as an installation.

The metadata I collected for each work was the name of the artist, the title and the date or period of creation. I excluded works for which one or more of these items was not available. The Metropolitan Museum of Art and MOMA provide the data on their collections in downloadable files.¹⁴ With the other institutions represented in my dataset, the data was downloaded directly from the online collections using bespoke computer scripts written in the open-source Python language.¹⁵ My dataset was constructed as a simple one-table Excel spreadsheet. I also used Python scripts to automate the repeated application of the statistical modelling tools I utilise in this chapter and to analyse the results of using those tools.

A list of the institutions whose collections are represented in my dataset is given in Table 5.1, which also gives the number of entries from each included in the dataset. As can be seen, there are 59,262 entries in total.

¹⁴ These are available through the GitHub open-source software and data repository, which can be found at <https://github.com/>. The Tate also provided this data through GitHub, but has not updated it since 2014 and it does not include a specification of the type of work. I queried the Tate's online collection directly to obtain the data required for my dataset.

¹⁵ Python can be downloaded from the website of the Python Software Foundation, which can be found at <https://www.python.org/>.

Art Museum	Entries
Albertinum Museum: Dresden, Germany.	1,141
Art Museum of the Americas: Washington DC, United States.	660
Art Institute of Chicago: Chicago, United States.	1,670
Castello di Rivoli, Museo d'Arte Contemporanea: Turin, Italy.	357
Centre Pompidou: Paris, France.	7,409
Daegu Art Museum: Daegu, Korea.	747
Daros Collection of Latin American Art: Zurich, Switzerland.	453
GAM (Galleria Civica di Arte Moderna e Contemporanea): Turin, Italy.	210
Guggenheim Foundation: New York, United States; Venice, Italy; and, Bilbao, Spain.	1,631
LACMA (Los Angeles County Museum of Art): Los Angeles, United States.	1,208
MAC (Museo de Arte Contemporaneo): Santiago, Chile.	456
MAC Lyon (Musée d'Art Contemporain de Lyon): Lyon, France.	635
MALBA (Museo de Arte LatinoAmericano de Buenos Aires): Buenos Aires, Argentina.	243
MALI (Museo de Arte de Lima): Lima, Peru.	213
MAM (Museu de Arte Moderna de São Paulo): São Paulo, Brazil.	768
MAMbo (Museo d'Arte Moderna, Bologna): Bologna, Italy.	821
The Metropolitan Museum of Art: New York, United States.	2,860
MMCA (National Museum of Modern and Contemporary Art): Gwaechon, Seoul, and Deoksugung, Korea.	2,393
MOCA (Museum of Contemporary Art): Tokyo, Japan.	1,188
Moderna Museet: Stockholm, Sweden.	5,222
MOMA (Museum of Modern Art): New York, United States.	3,850
Musée d'Art Moderne de la Ville de Paris: Paris, France.	2,862
Musée d'Orsay: Paris, France.	1,567
Museo Coleção Berardo: Lisbon, Portugal.	921
Museo Naçional de Bella Artes de Cuba: Havana, Cuba.	403
Museo Novecento: Florence, Italy.	189
Museo Naçional Centro De Art Reina Sofia: Madrid, Spain.	2,472

Table 5.1. Art Institutions in my dataset.

Museum of Fine Arts - Latin American Art Collection: Houston, United States.	291
National Museum of Art: Osaka, Japan.	639
National Museum of Modern Art: Kyoto, Japan.	1,667
National Museum of Modern Art: Tokyo, Japan.	1,811
Pinakothek Der Moderne and Museum Brandhorst: Munich, Germany.	3,310
Städel Museum: Frankfurt, Germany.	511
Stedelijk Museum Amsterdam: Amsterdam, the Netherlands.	2,753
Tate: London, Liverpool, and Saint Ives, United Kingdom.	5,731
Total all institutions	59,262

Table 5.1. Continued.

The interpretations I offer in the remainder of this chapter are therefore of modern and contemporary art as it has been represented through the collecting policies and practices of the institutions listed in Table 5.1. Over and above the geographical constraints, the selection of period, and the choice of media I have already discussed, my reading will also reflect to some extent the inclusions or exclusions of certain types of artist or work in those collections. Several of the collections are heavily weighted towards artists in the canonical history of modernism, and some include a large number of works left to the institution by artists in their estates.¹⁶ The art historian Terry Smith has argued that 'no museum has succeeded fully in becoming a site for contemporary art', and 'in general ... the leading centers in Euro-America celebrate those artists who ... perpetuate traditional subjects'.¹⁷ Women are substantially under-represented in the collections included in my dataset, with that bias being most acute with artists active in the early-twentieth century.¹⁸ I have not made any adjustments to the data to reflect these factors as there is no principled way on which to make them. In addition, as

¹⁶ One measure of how concentrated my dataset is with a small number of artists is that over 50% of the artists included in my dataset have only one work represented, whereas 1% have 50 or more, and together account for 18% of all works.

¹⁷ Terry Smith, *Contemporary Art: World Currents*, (London: Laurence King, 2011), p. 45.

¹⁸ See for instance Cornelia Butler and Alexandra Schwartz, (eds.), *Modern Women: Women Artists at the Museum of Modern Art*, (New York: MOMA, 2012), and references therein. My dataset gives some idea of the level of under-representation through comparison with contemporary exhibition catalogues. In all decades from the 1900s to the 1940s works by female artists represent around 5% of the works in my dataset to which a gender was assigned. Since then, the proportion has grown steadily to around 30% in the 2000s. If we look at the French art museums in my dataset, women were responsible for 4% of the works made in the 1900s and 7% of the works created in the 1920s. In comparison, exhibition catalogues indicate that women produced 15% of paintings on show at the Paris Salon d'Automne in 1904, and 18% of paintings on display at the Paris Salon des Indépendants in 1928. Société du Salon d'Automne, *Catalogue de Peinture, Dessin, Sculpture, Gravure Architecture et Arts Decoratifs*, (Paris, 1904), and Société des "Artistes Indépendants", *Catalogue de la 39^e Expositions*, (Paris, 1928).

will be seen, my modelling allows for the impact of some to be isolated, and for others the underlying trends are still apparent although with greater volatility.

With most items for most of the art museums some additional data processing was required, and this was also carried out through Python scripts with some additional manual processing. I removed any duplicates as identified from their collection numbers, for instance the Centre Pompidou's online collection can include multiple records for the same object if the institution has more than one image of the work. In some cases, such as in the MALBA online collection, titles may be presented with an English translation in brackets. In others, brackets may be used to hold additional information on the work of art provided by the museum, for instance giving the purpose for which the work was originally commissioned. Brackets were also used by artists themselves. Several of the artists I have modelled in my second case study included bracketed text in some of their titles. Given these different uses of brackets, manual processing was required to distinguish them, and I removed translations and other curatorial additions from the dataset where I was confident the bracketed text had been used for that purpose. However, it is very likely I have not identified all instances of curatorial additions in brackets. I also tidied up the dataset in ways affecting only a very small number of entries. For instance, some entries related to canvases with paintings on both sides and gave two titles - one for the recto and one for the verso. In these cases I retained the title for the recto.

To look at the patterns in language use across the titles in my dataset I translated them all into English. Most of the statistical analyses I made of my whole dataset would not have been possible without this step and I would have been restricted to comparisons of separate analyses for each of the languages used. It was also the most effective way of including in my analysis titles in languages such as Portuguese that represented only a small proportion of my dataset and where the results of analyses looking at that language alone would have been much less robust. In addition to art museums in English-speaking countries, several of the museums in my dataset provide titles in English. For those which do not, I utilised the Google Translate service to translate from the native language of the country in which the museum was based.¹⁹ Instances of non-native and non-English titles such as a German title in the Museo Coleção Berardo's online collection were identified manually, consolidated and processed through Google Translate.²⁰ This was supplemented by manual examination and automated replacement of common mistranslations or multiple translations. For instance, the

¹⁹ The Google Translate Service can be found at <https://translate.google.co.uk/>.

²⁰ I investigated the use of automated language recognition to support this task but found that the error rates were too high to make the process more efficient than one involving manual examination only.

French word 'Nu' was often translated as 'Naked' when 'Nude' was the appropriate translation. One example of multiple translation is that in French museum collections with multiple versions of Josef Albers' *Hommage au Carré*, Google Translate would alternately render the titles in English as *Homage to the Square* or *Tribute to the Square* and all instances of the latter were transformed to the former as it is the title used in English-language museums.²¹ In a final step I identified the most common variations in spelling between American English and British English such as 'color' and 'colour', and transformed all instances of American English to British English. Google Translate provides literal translations with at times an unnatural word order in English and my dataset is likely to include a residual number of mistranslations and spelling variations. Notwithstanding these shortcomings, the English translations in my dataset are adequate for the types of reading of titles I perform in this chapter where titles are looked at in aggregate and large-scale trends are given art historical interpretations. It is not my aim to provide close readings of the kind developed by Christine Mehring of Gerhard Richter's *Abstraktes Bild* which I discuss in Chapter 2, where the appropriate translation of 'bild' is a key part of the interpretation.²²

I made further transformations of the data for inclusion in my dataset. In titles there are several ways of indicating a numbered work, including 'number', 'no.' and '#'. I replaced all of these with 'number', to allow the overall trends in the numbering of works to be identified. In a manual pass I transformed all Roman numerals to Arabic numerals, to allow the levels of use of each particular number to be identified and to avoid any ambiguity between the Roman numeral 'i' and the first-person singular pronoun 'I'. All artists' names were processed to be in the format 'given name family name' and all accents were removed, to allow for sorting and identification of all works by the same artist. For each work I recorded a single year as that in which it was created, to enable an analysis of trends over time. In the large majority of cases the online collections record a year, a season, or a date in a year on which the work was created, and that year was entered into the dataset. If the date was given as 'around', 'before' or 'after' a certain year I recorded that year. In some cases, the online collection gave a range of years. I excluded items where the ranges were longer than 20 years, otherwise I took the middle of the range as the year of creation. This will have introduced a bias into my dataset, boosting the number of works recorded as being created in those middle years, but will not have

Google Translate is based upon a non-deterministic algorithm, which is responsible for the multiple translations of repeated titles such as *Hommage aux Carrés*.

²² Christine Mehring, (2012), pp. 21 - 22.

affected the long-run trends. Table 5.2 gives the number of entries in my dataset with a given or estimated creation date in each decade from the 1900s to the 2000s.

Decade	Entries
1900s	2,393
1910s	4,045
1920s.	5,212
1930s	5,686
1940s	4,488
1950s	6,056
1960s	9,017
1970s	6,314
1980s	6,769
1990s	5,064
2000s	4,218
Total all decades	59,262

Table 5.2. Distribution of titles by decade, 1900s to 2000s.

The total number of artists with works in my dataset is around 15,500 and Table 5.3 gives the number represented in each decade (with an artist creating works in more than one decade counted once for each).

Decade	Artists
1900s	969
1910s	1,219
1920s	1,824
1930s	2,132
1940s	1,846
1950s	2,318
1960s	3,487
1970s	2,945
1980s	3,145
1990s	2,355
2000s	2,061

Table 5.3. Number of artists with works created in each decade, 1900s to 2000s.

To classify each artist as male or female I first matched their name against that in the Github data provided by the Tate and MOMA, both of which give an artist's gender. For the artists who do not appear in those collections I matched their given name against given names taken from the Tate and MOMA data and lists of the 100 most common male and female given names in Dutch, French, German, Italian, Portuguese, Spanish and Swedish. There is no

standard way Japanese and Korean names are rendered in the Roman alphabet, and there were spelling variations in the names across the collections in my dataset. In addition, unisex given names are not uncommon in Korean. For these reasons I made a manual pass through the Korean and Japanese names in which I identified possible instances of multiple names and rendered them all as the most commonly occurring where I was confident that was the case. I then took the most frequently occurring artists and looked to classify them as male or female through researching them online. My data on given names also enabled me to identify those that can be male and female in the same or different European languages, such as Francis, Jean, and Joan and in another manual pass through the data I looked to correct any mis-assignments. Overall, I classified the artist as male or female with 56,676 entries, or 95% of the total. My dataset will include some gender misclassifications, and artists who would not identify as male or female. It is also likely that there are multiple names for the same Korean or Japanese artist remaining in my dataset. However, these misclassifications will be a small minority of entries.

My analysis of trends over time in the language used in titles is complicated by the fact that in some cases the titles as they are now presented would not have been part of the public face of titling at the time the works they name were created. The titles of works may have been changed after they were first exhibited, either by the artist, an owner, or a dealer. In other cases, works may not have been put on public display when created, for instance remaining with the artist until their death. Mark Rothko is an example of both. He sometimes re-titled works and many of his paintings in the collections included in my dataset came from his estate and were titled by the gallery to which they were turned over by his executors.²³ With some works the original title may have been lost and the work recorded as 'no title' or 'untitled'.

Without a detailed study of the history of a sufficiently large selection of the titles in my dataset it is not possible to say with a high degree of certainty how the readings I present in this chapter have been influenced by these factors. However, I have reviewed the *Catalogue Raisonné*s of five artists and this gives some comfort the impact is not substantial. In most cases of re-titling or of titles given to works left in estates, the titles are of the same kind as the titles of other works displayed by the artist at around the time they were created. For instance, Mark Rothko often re-titled by re-numbering paintings, and the titles used by the gallery managing his estate mainly consisted of colour words, a practice Rothko also followed with some of his paintings. In addition, works where the original title has been lost are most likely to date from

²³ For more detail see David Anfam, *Mark Rothko: The Works on Canvas* (New Haven and London: Yale University Press, 1998), pp. 5 - 21.

the early twentieth century. Works from that time recorded as 'no title' or 'untitled' make up a very small proportion of my dataset.

5.3 Titling in Modern and Contemporary Art

Correspondence Analysis - Linguistic Variation

The first vantage point for my interpretation of what titles in aggregate can tell us about the history of modern and contemporary art is that provided by correspondence analysis. This produces a compact visual overview of my data which enables relationships between the words used in titles and the years in which the works they name were created to be read off. An implementation of correspondence analysis is available as part of the Voyant web-based text analysis environment developed by the scholars Geoffrey Rockwell and Stéfán Sinclair.²⁴ I used this tool with the English-language translations of the titles in my dataset, split by decade from the 1900s to the 2000s. Combining the titles by decade is an appropriate way of investigating long-term trends in the use of the language in them. As I wanted to look at all linguistic components as of potential interpretive significance I processed the titles to identify the main punctuation marks and other symbols such as commas, quotation marks and exclamation marks and replaced them with identifying texts to make them more visible on the chart.²⁵ The results are presented in Chart 5.1, which looks at the eleven decades and the 122 most frequently-occurring words or other linguistic components.²⁶

The mathematics of correspondence analysis dictates where the word and decade labels are located on the chart and the Voyant tool does not give any choice to the user in the colours that are used. As a consequence, Chart 5.1 is more complex and harder to read than the other charts I present in this thesis, and I will now give a detailed explanation of what the chart contains and how to interpret it. Each of the words is presented in black type face above a coloured circle, and the size of the circle associated with each word represents its frequency.

²⁴ The Voyant text-analysis environment can be found at <https://voyant-tools.org/>. For an application of correspondence analysis to the humanistic interpretation of texts see Geoffrey Rockwell and Stéfán Sinclair, *Hermeneutica: Computer-Assisted Interpretation in the Humanities*, (Cambridge: The MIT Press, 2016), pp. 69 - 82.

²⁵ The identifying texts were chosen so they did not feature in any of the titles in my dataset. Those appearing in the chart are as follows: 'lb' left bracket, 'rb' right bracket, 'pperiod' period, 'comma' comma, and 'quotes' quotation mark.

²⁶ The user of the Voyant correspondence analysis tool can choose the number of most frequent words to be displayed. As all words with the same frequency are presented on the chart, the actual number of words displayed may be higher than the number selected by the user if there are several with the same frequency. This is the case with Chart 5.1, where I selected 100 as the number of most frequent words to be displayed, but where frequency ties meant that 122 are given in the actual plot.

The decades are in pale blue, above a lozenge of the same colour. The two dimensions presented on the chart are those along which the data varies the most, in this case they capture 84% of the variation and so it is unlikely any major dependencies have been missed.

The Voyant tool allows the words to be clustered in different ways and visual examination of the chart indicated three clusters worked well, which are coloured blue, pink and green in the circles below each word in Chart 5.1. As can be seen, the words in each are of distinct types. The blue cluster includes numerals and punctuation marks, as well as the words 'untitled' and 'series'. The green cluster has colour words, formal words such as 'painting', 'sculpture' and 'composition' and the term 'number'. The pink cluster includes nouns such as 'woman' and 'flowers', genre labels such as 'portrait' and 'still-life' and the words 'the', 'by' and 'of'. Voyant does not allow the decades to be clustered, but it is apparent that they also group together, with the early twentieth century in one group in the top left, those from the late twentieth and twenty-first century in another on the top right, and the 1950s and 1960s in a third. The 1940s and 1970s sit between clusters.

Interpretation of the chart is guided by a number of principles. The further a word or decade is from the point labelled (0,0), or the 'origin', the more discriminating it is in that its profile of use differs the most from the average.²⁷ Chart 5.1 indicates that of the 122 most frequent words, use of 'homage' and 'composition' varied most between decades. In addition, the closer two words are the more similar their patterns of use across decades. Likewise, the closer two decades are to each other the more similar the profile of words used in titles during those periods. Chart 5.1 shows that the profiles of the words used in the titles of works created in the 1990s and 2000s are very similar, and are most dissimilar to the words used in titles during the 1900s.²⁸

The closeness of words and decades in the chart also give an indication of how the usage of words in titles changed over time.²⁹ Words and decades on the same side of the origin are likely to have a positive association, in that the words were used more heavily in those

²⁷ Given the importance of position relative to the origin in interpreting the results, one shortcoming of the Voyant correspondence analysis tool is that it does not present vertical and horizontal axes going through the origin. I have added these axes by hand to Chart 5.1.

²⁸ The scaling of the axes on the chart comes from the way profiles are represented in the mathematics of correspondence analysis. The closeness of words or decades has, as I have discussed, a meaning. However, the sign of the co-ordinates on each dimension does not. For instance, a chart in which the left-hand quadrants and right-hand quadrants are mirrored provides the same information as Chart 5.1.

²⁹ As I explain in Appendix C, the conclusions that can be drawn on the relationships between word and decade are less robust than those between words themselves or between decades themselves, and may be misleading. Another shortcoming of the Voyant website is that it is not explicit on this potential risk to users in interpreting its graphical presentation of the results of correspondence analysis.

decades than in others, whereas those on opposite sides are likely to have a negative association. We can see that it is likely there was a strong trend in the use of words as you move through the twentieth century. Starting at the top left of the chart, the words in the pink cluster were the most heavily used in the early decades of the twentieth century, with relatively little use in titles at that time of words in the other clusters. As you move down the chart and forward from the 1900s the strength of the association weakens, and the words from the green cluster begin to appear more often in the titles of each decade. In the 1940s, the words in both the green and pink clusters are appearing at significant levels (most of the words in both clusters are on the same side of the origin as that decade, whose label is partly obscured by the word 'window'). By the 1950s the words in the green cluster feature most heavily, and at this time significant levels of use of words in the blue cluster are beginning. As you move back up the chart towards the top right and forward in time from the 1960s to the 2000s, the use of words in the blue cluster grows, and of words in the green cluster declines.

As my use of the technique shows, correspondence analysis is a powerful and compact way to identify long-run trends and other patterns in the use of words in the titles. However, it has some significant shortcomings, which means it would be premature of me to look to give an art historical interpretation of the trends at this stage. The chart is visually very compressed, and it is not feasible to examine more than around 100 words in this way on a single chart. To see if there are any significant associations involving less frequent words would require additional charts to be generated. The clustering of words also needs to be interpreted carefully to avoid misreading the chart. For instance, 'number' is in the green cluster and the numerals 1 to 7 in the blue cluster. This does not mean the words were not used together, but that numerals were used on their own more in the decades from the 1980s onwards than in the 1950s to the 1970s, whereas they were used more often in association with 'number' in those earlier decades. The nature of correspondence analysis means what can be derived directly from the chart is largely comparative and qualitative, involving the relative strength of associations and other dependencies. Finally, as I have discussed, only provisional conclusions can be drawn about the inter-relationships of word and decade. To develop a more comprehensive and quantified understanding of the trends and clusters revealed by the correspondence analysis and to give an art historical interpretation to them, I turned to the natural language processing techniques of topic modelling and parts of speech tagging.

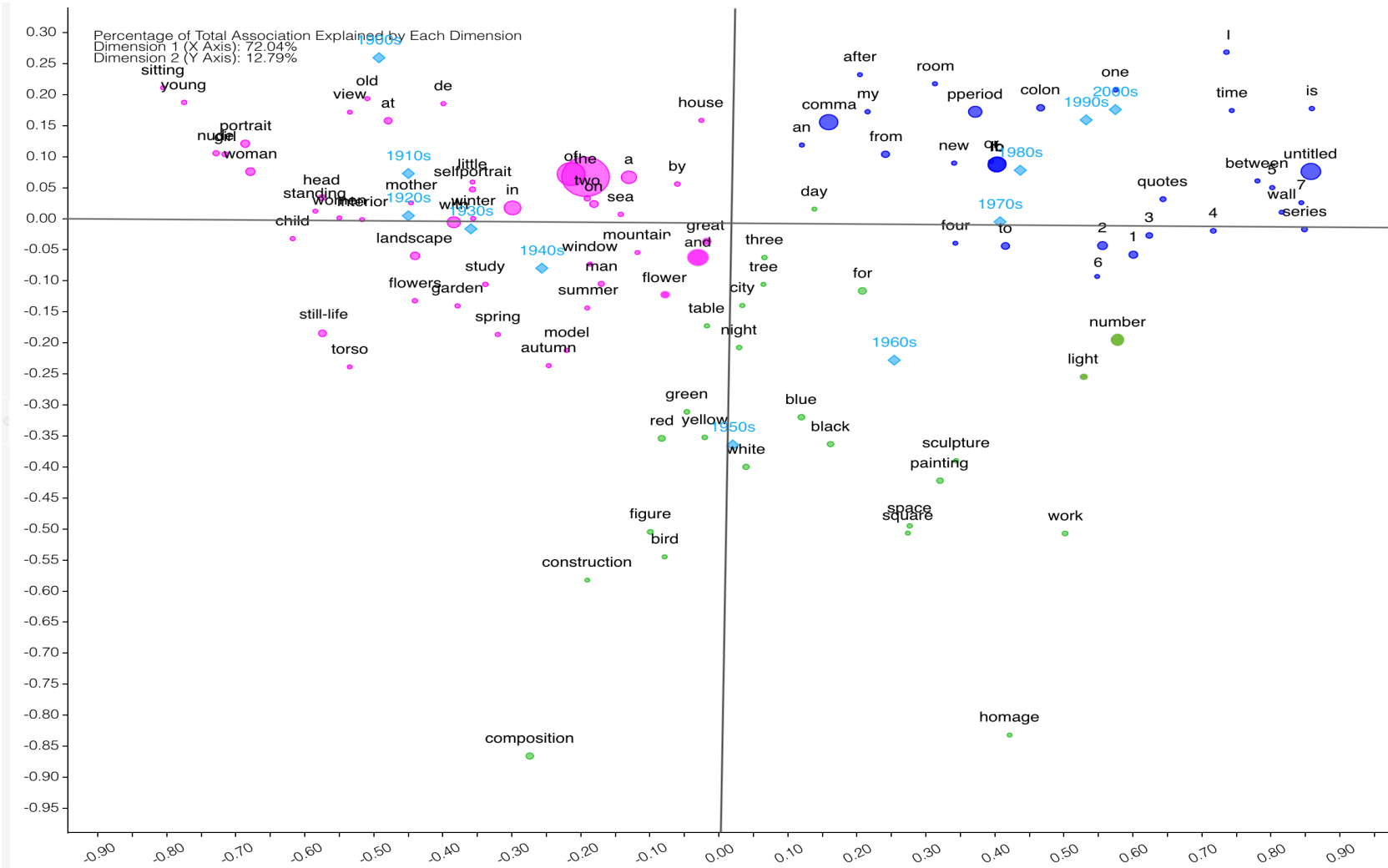


Chart 5.1. Correspondence analysis presenting the associations between the 122 most frequently occurring words and other linguistic components and the decades in which they were used.

Topic Modelling - Titles as Signals of Artistic Interests

Topic modelling is a predictive technique in which a statistical model is developed of a collection of documents.³⁰ Each 'topic' is a weighted mix of all the words occurring across those documents, and each document is modelled as a weighted mix of all the topics. Prominent topics which feature with high weight across several documents represent recurring patterns of use of the words of highest weight in those topics. Topics prominent only in a single text or a small number of texts contain words whose use is highest in those documents. Topic modelling is therefore a useful tool for the identification of clusters and other patterns of language use across a body of texts. However, they also present interpretive challenges to the humanities scholar. In topic modelling each word is treated merely as a string of characters. For instance, a topic model counts two instances of 'bank' in a text as occurrences of the same word, even if in the first case it means a financial institution and in the second the side of a river. Topic models also follow a 'bag of words' approach in which the order of the words in the documents being modelled is not considered.

The implementation of topic modelling I utilised is that developed by Andrew McCallum as part of his MALLET toolkit, and in what follows I describe topic modelling of the titles in my dataset using that tool.³¹ Other implementations of topic modelling take a similar approach. I began by excluding what are known as 'function' words from my analysis. These have little meaning on their own and their role is to create grammatical or structural relationships between the 'content' words, which have an independent meaning. Function words include the conjunctions (such as 'as', 'and', and 'so') and the definite and indefinite articles. They are very common in titles and so would dominate the topics in the model if retained, making those topics harder to interpret. I also excluded all punctuation marks and other non-alphanumeric symbols except for the hyphen, and transformed all words to lower case. The hyphen is commonly used in titles such as 'self-portrait', and to have excluded it would have split such titles into two separate words each of which would have been treated independently in the

³⁰ Topic modelling was developed by computer scientists in the 2000s, see David M. Blei, Andrew Ng and Michael Jordan, 'Latent Dirichlet Allocation' in *Journal of Machine Learning Research*, 3, (2003), pp. 993 - 1,022. In 2012, an entire edition of the *Journal of Digital Humanities* was devoted to topic modelling, which can be found at <http://journalofdigitalhumanities.org/2-1/>, and a 2013 edition of *Poetics* looked at their application in the cultural sciences, which can be found at <https://www.sciencedirect.com/journal/poetics/vol/41/issue/6>. For a technical introduction and a survey of their use across a range of disciplines see Jordan Boyd-Graber, Yuening Hu and David Mimno, 'Applications of Topic Models', in *Foundations and Trends in Information Retrieval*, XX/XX, (2017), pp. 1 - 154.

³¹ Andrew Kachites McCallum, 'MALLET: A Machine Learning for Language Toolkit.', which can be found at <https://mimno.github.io/Mallet/index>.

modelling. I then experimented with varying the various inputs and parameters available to the researcher with the MALLET topic modelling tool, such as the ways in which titles are grouped together and the number of topics to be identified. My aim was to see if a model could be developed that supported an art historical interpretation utilising as much as possible of the information contained in the model. The topic model meeting those aims looked at the titles used in each decade from the 1900s to the 2000s and modelled them as being generated by 20 topics. Models that looked at each title as a separate text or grouped them together by year were not as well-suited for the identification of long-term trends. Models with fewer than 20 topics did not discriminate between the words used in different decades as well as the 20-topic model, and those with more included topics that were redundant in having low weight across all decades.

A common approach to the interpretation of a topic model is to identify and focus on those topics where the most important words seem to coalesce around a theme. In this approach topics are treated univocally, having a stable meaning across the documents being modelled, and may be considered in isolation. As I discuss in the literature review, Matthew Jockers follows this approach in his topic modelling of a corpus of nineteenth-century novels.³² A key interpretive issue with thematic readings is that in treating them as strings of characters, the topic model knows nothing of the meaning of the words in each topic and how that might have changed over time. And, indeed, Jockers is mindful of this issue. He does not take the consistency and univocality of his topics for granted but argues they can be interpreted that way.³³ This is not the only approach to developing and reading a topic model. Topic models are rich sources of information on the documents they model and, in focusing solely on the theme, thematic readings often do not look at the differences in weight or the associations between the important words in topics. Topic models not supporting a thematic reading can also be of value to the humanities scholar. As I also discuss in the literature review, the literary historians Andrew Goldstone and Ted Underwood have investigated other ways of using topic models in humanistic scholarship.³⁴ For instance, they read changes in sense of a particular word from the associations between it and the other important words in different topics prominent in the model for different texts.

³² Jockers, (2013), pp. 118 - 153.

³³ Jockers, (2013), p. 132.

³⁴ Goldstone and Underwood, (2014). For another scholar who has taken a similar approach to the interpretation of topic models see Piper, (2018), pp. 66 - 93.

As will be seen, my topic model can be given the kind of rich interpretation developed by Goldstone and Underwood. The ways words re-appear across topics, the importance of words within topics, and the prominence of topics in a text and across texts are all of interpretive significance. In my reading I will return to the three basic functions of the title I discuss in the introduction.³⁵ These are the ‘nominative’ function of naming the work, the ‘semantic’ function of acting as a commentary and contributing to the meanings it is given, and the ‘seductive’ function of attracting the attention of the reader.

My model is dominated by three topics, whose most important words correspond closely to the content words in the three clusters of the correspondence analysis. Looking at how the importance of these topics changes by decade confirms and gives a quantified version of the strong associations between word and decade I set out in my reading of the results of that analysis. For every decade from the 1900s to the 2000s the most prominent topic in the model is one of those three topics. As can be seen from Chart 5.2, their combined weight within a decade is never less than 60%. The first, ‘topic 1’, dominates in the early decades of the twentieth century; ‘topic 2’ rises to be the most prominent in the 1950s and 1960s before declining, and ‘topic 3’ is the most important in my model for the decades from the 1970s to the 2000s. No other topic is prominent in the model for more than one decade.

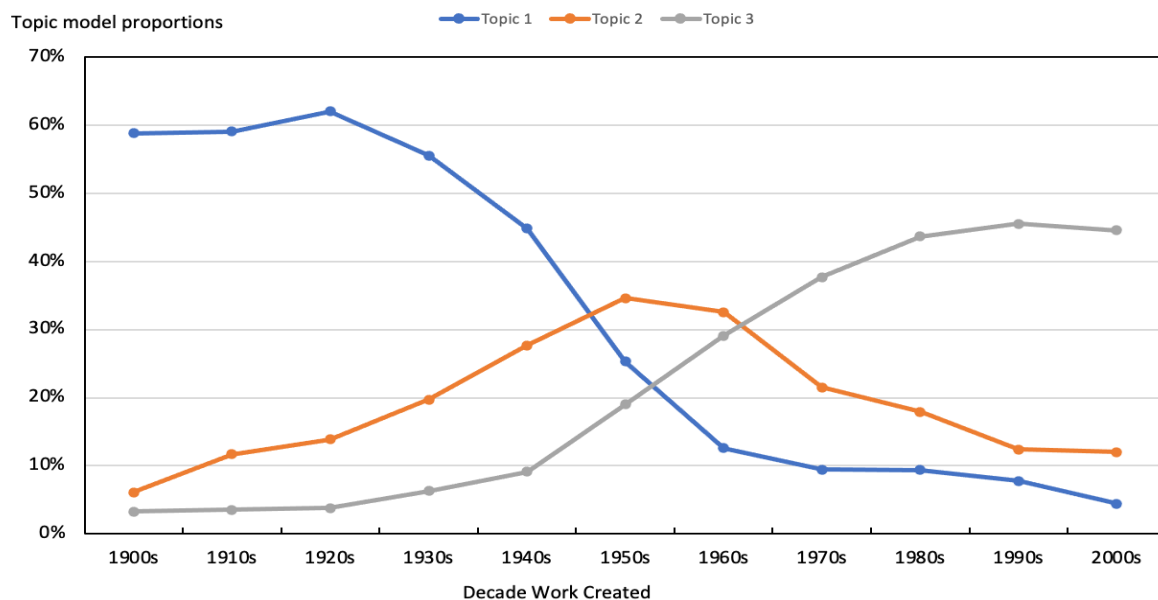


Chart 5.2. The three dominant long-term topics in the 20-topic model, weights in the model for each decade, 1900s to 2000s.

³⁵ Genette, (1997), and Besa Camprubi, (2002).

To give an interpretation of a topic you look at its most important words. A succinct way to do this is through the word cloud, which presents the frequencies of the words in a text in a manner in which that information can be readily assimilated.³⁶ In my case I generated the word clouds with frequencies given by the weights of the 50 most important words in each of topics 1, 2 and 3. In each word cloud, it is only the relative size of words that is of interpretive significance - the larger the word the higher the weight it has in the topic. The relative position and orientation in the cloud are not important – two words being close together or having the same orientation does not indicate any relationship between them. The cut-off at 50 words did not represent a sharp break within each topic, with the weights only diminishing slowly at that point as you move through the words of a topic in descending order of weight. I chose it as a cut-off at 25 words excluded some of interpretational significance whereas cut-offs at 75 and 100 words did not add anything significant to my interpretation. All the words bar the first person singular 'I' are in lower case. The basic meaning of the words in the clouds for topics 1, 2 and 3 will not have changed over the course of the twentieth century, although, as we will see, some of their associations have.

The word cloud for topic 1 is presented in Figure 5.1. The most important words are predominantly those relating to the generic categories of landscape, interior, portrait, and still-life or to the most common types of figurative subjects including the seasons of the year, gardens, nudes, children, and flowers. Used in titles they would most likely have expressed some sort of artistic engagement with the figurative tradition or with figurative subject matter. As a shorthand, we can characterise topic 1 as 'figurational'. Chart 5.2 shows that these figurational interests were the most prominent signalled through the language used in titles during the first five decades of the twentieth century.

³⁶ Word clouds have come under heavy criticism in recent year for being over-used, reductive, and potentially misleading, as Jacob Harris amongst others has argued in his blog which can be found at <http://www.niemanlab.org/2011/10/word-clouds-considered-harmful/>. However, they remain powerful tools when used appropriately and interpreted properly. In my case they allow the rapid visual appreciation of the relative importance of the words in a topic.

in the top 50. The words 'abstract' and 'form' are also important in this topic. Colour words were used in the titles of figurative paintings, however only one - 'red' - appears in the top 50 most important words with topic 1. Their stronger presence in topic 2 shows their use was increasingly dissociated from figurative subject matter and so can be read as expressions of an interest in colour itself as a formal element of a work.

'Number' is the word of highest weight in topic 2 and the numerals 1, 2 and 3 also feature in the top 50. Numerals can function to identify or to order. However, the 'bag of words' approach with topic modelling does not allow these uses to be distinguished as it does not look at how words occur together in titles. The title 'Number 1' might signal a desire to minimise or eliminate the title's semantic role as something contributing to the work's meaning, restricting its function to the nominative one of identification and nothing more. In contrast, the words 'Number 1' in the title 'Compositions in Red and Blue Number 1' might indicate a work to be seen as the first of a sequence and so function as a form of commentary. To understand how the word 'number' and numerals functioned we need to look at their presence in whole titles, not just as individual word tokens. I will come on to do this when I look at my parts-of-speech analysis and at some descriptive count-based statistics.

top 50. With topic 2 they would appear to indicate a stronger association with ordering or numbering individual works than with works to be seen as part of a larger whole. Indeed, 'series' and 'part' do not appear in the top 50 words for topic 2, whereas 'number' is more prominent in topic 2 than in topic 3. As with the change in the associations of colour words and of 'composition' from topic 1 to topic 2, we can read the topic model as showing how artistic interests have persisted but at the same time become re-focused or re-inflected.

The other important words of topic 3 can be read as signals of a range of possible artistic interests. Words such as 'art', 'museum', 'room', and 'wall' might indicate an engagement with the institutional context within which work of arts are displayed. In association with these interests, the words 'painting' and 'work', which carry over as important words from topic 2 to topic 3, might indicate a re-inflection of interest away from the work of art itself towards its institutional context. The occurrence of 'I', 'love', 'death', 'eyes', 'mirror' and 'memory' suggest an engagement in some way with the somatic or the personal. 'Self-portrait' features as an important word in topic 3 and topic 1, but not topic 2, and when seen alongside the other important words in topics 1 and 3 suggests a shifting of interests in self-representation away from the figurative towards the personal. However the important words in topic 3 are read, and other readings are clearly possible, the interests expressed by them are more heterogeneous than those in the other two. I have characterised topic 1 as 'figurational' and topic 2 as 'abstract/formal'. Topic 3 does not have such a summary.

To complete my reading of the three most prominent topics in my model, we can note that there are no words that are important in all of them, which is an indication of how the strongly the interests expressed through titles changed over the course of the twentieth century.

We can also look at the art historical significance of the other topics in my model. In almost all cases they feature as a prominent topic in one decade only. Typically, they can be read as reinforcing the interpretations I have given of topics 1 to 3. Often, they include words particularising the generic language featuring in those topics. From the 1900s to the 1940s, the topics prominent in those decades include important words giving specific locations or types of object as well as the first names of individuals. Signals of artistic allegiance also feature, with 'cubist', 'suprematist' and 'dada' words of high weight in prominent topics from that period. The practice of numbering works can be seen in the numerals of years featuring as important words in topics prominent during each of the decades from the 1950s to the 2000s.

The theme of artistic engagement with the idea of composition that John Welchman traces out in his history of titling comes out strongly in my topic model.³⁷ As I have already discussed, its appearance as an important word in topics 1 and 2 shows how that engagement was re-inflected over time. It also has by far the highest weight in the topic prominent only in the 1950s, indicating a peak in direct artistic engagement with the compositional in that decade over and above the persisting associations captured in topic 1 and topic 2.

The other important words in these topics often relate to particular subjects of artists heavily represented in the collections included in my dataset. In isolating these words in particular topics, the model is partly correcting for the biases introduced into my dataset by these artists. The topic prominent for the 1940s includes words reflecting the type of subject popular with the artist George Roualt, whose has over 300 works from that period in the Centre Pompidou's collection. From the 1960s to the 2000s the important words include those relating to multiples or series of works by individual artists. The presence of works by Andy Warhol from the 1960s and 1970s in many of the collections is reflected in the words 'Mao' and 'Brillo' being amongst the most important in topics prominent during those decades. Other artists who produced works in series include Cildo Meireles with his *Insertions into Ideological Circuits*, which appear in multiple collections, and Kiki Smith's series *Jersey Crows*. The words from these titles are important with topics of high weight for the decades during which the works were produced.

The art historical value of the topic model is not limited to that provided by the most prominent topics in each decade and the most important words in those topics. The overall distributions of topics for each of the decades from the 1900s to the 2000s can give a broad idea of the diversity of titling practice in each period and how that has changed over time. The more diverse the titles used by artists the more topics there will be of significant weight in the model for that period to capture that variety. Conversely, the more titling is centred upon a limited set of words, the more the topic model will be concentrated in a few topics of high weight. A succinct measure of diversity or complexity in this sense of the degree of concentration of the weights in a topic distribution is its 'entropy'. Entropy can be thought of as a generalisation of the number of 'bits' of information required to characterise the distribution.³⁸ As I discuss in

³⁷ Welchman, (1997).

³⁸ To illustrate how entropy is calculated we can look at two hypothetical examples of topic models for the titles used in a particular decade. In the first example the titles are generated exclusively by 2 topics of equal weight. In this case the entropy of the topic distribution is 1 bit (a bit can take on one of two values, 0 or 1, corresponding to the two topics that make up the distribution). In the second example the titles are generated equally by 16 topics. In this case the entropy is 4 bits (in binary arithmetic there are 16 binary numbers of 4 digits). For a mathematical introduction to entropy see Thomas M. Cover

the literature review, scholars have used the concept to investigate the diversity of production of works in different genres by Dutch artists.³⁹ Other scholars have used it to look at questions of cultural or linguistic complexity.⁴⁰

The entropy of the topic distribution in my model for each decade is shown in Chart 5.3. From an art historical perspective, the absolute levels of the entropy are less important than the change in its value over time. As can be seen, on this measure titling became persistently and significantly more diverse from the 1920s to the 1980s. As artists moved away from signalling a predominant interest in, or engagement with, figuration, in aggregate the words in the titles they used in each decade were sending out a wider range of signals. The 1950s to the 1980s was a time of considerable artistic flux during which topics 1 and 2 declined significantly in importance in my model, while topic 3 rose to be the most important. From the 1980s onwards the diversity of interests signalled through the titles used in each decade has changed little, with the entropy of the topic distribution roughly constant.

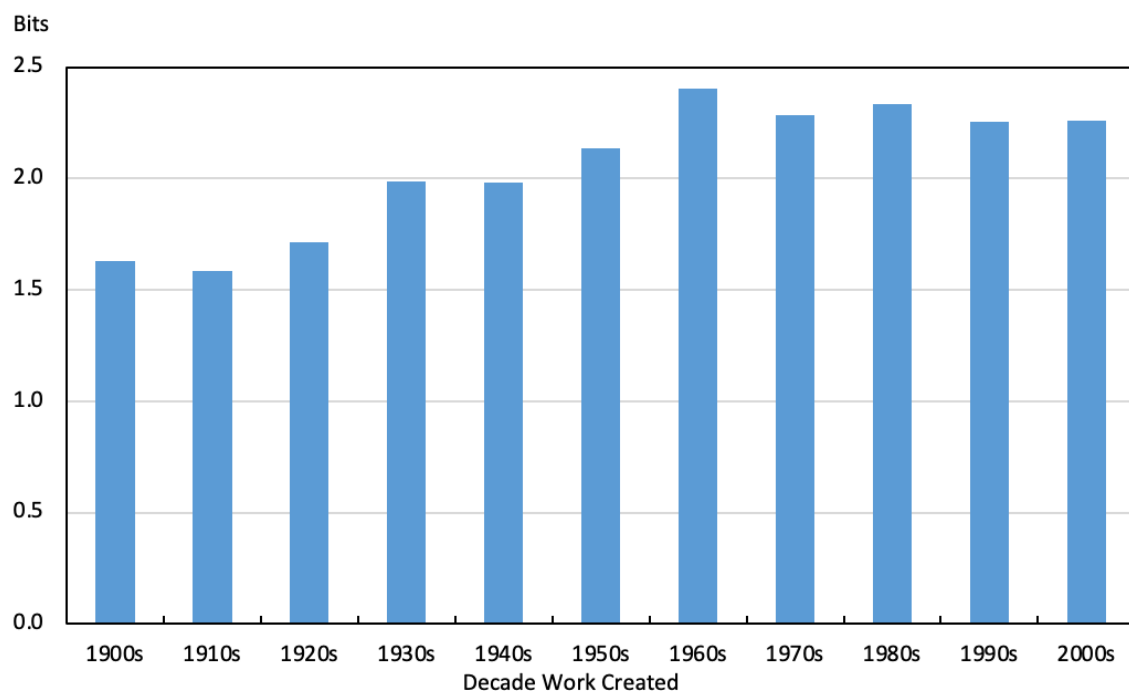


Chart 5.3. Entropy of the topic model distributions for each decade, 1900s to 2000s.

and Joy A. Thomas, *Elements of Information Theory*, (New York: John Wiley & Sons Inc., 1991), pp. 12 - 49.

³⁹ Lincoln, (2016b), and Rasterhoff, Beelen, Li, and Kisjes, (2018).

⁴⁰ Juola, P. 'Using the Google N-Gram Corpus to measure cultural complexity', *Literary and Linguistic Computing*, 28(4), (2013), pp. 668 -675, and Ruina Chen, Haitao Lui, Gabriel Altmann, 'Entropy in Different Text Types', *Digital Scholarship in the Humanities*, 32/3, (2017), pp. 528 - 542,

A much more widely used and more readily understood measure of linguistic variety in a given collection of texts is the type-token ratio, and I was interested to compare these two measures.⁴¹ The higher the type-token ratio the less often words are re-used and so the greater the diversity of the words used in a text. I took as my base the decade with the fewest words, the 1900s, and for each subsequent decade selected the same number of words from a random selection of all the titles from that period. The results of my analysis are presented in Chart 5.4. As can be seen, it largely confirms the trends in the diversity of titling practice as measured by the entropy of the topic model distribution for each decade. Titles are more diverse in the second half of the twentieth century and the early twenty-first than in the first half of the twentieth. In titles for the decades from the 1900s to the 1950s each word type was used around three times on average. From the 1960s each word type featured around two and a half times on average in the titles used in each decade.

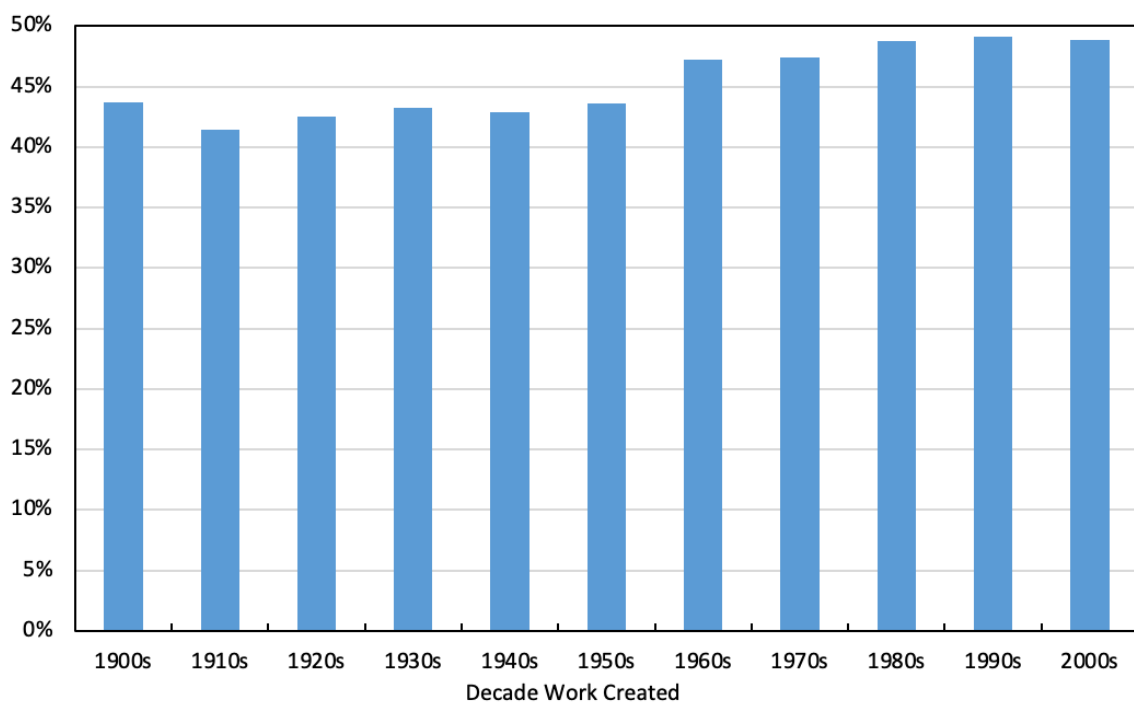


Chart 5.4. Type-Token ratio for titles in each decade, 1900s to 2000s.

⁴¹ The total number of word tokens in a text is the word count, and the total number of types is the number of distinct words. For instance, 'the cat ate the mouse' has five tokens and four types as 'the' occurs twice and the three other words once, and so has a type-token ratio of 4/5 or 80%. Type-token ratios can only be compared meaningfully between texts with the same word count. For an application of the type-token ratio as a measure of the diversity of literary titles see Jockers, (2012), pp. 54 - 58.

Parts of Speech and Count-based Statistics - Engagement with the functions of the title and signals of epistemic perspectives

My topic modelling delivers an understanding of the large-scale trends in the use of the 'content' words of titles during the twentieth century. The use of other components of titles may also be of art historical significance. The predictive technique of parts-of-speech tagging is one way of investigating that question, as well as other issues more directly related to my interpretation of the topic model. It allows the use of all classes of 'syntactic token' including different types of word such as nouns, adjectives or verbs, and other types of token such as numerals, symbols, and punctuation marks to be measured. In my analysis I used the state-of-the-art tagger developed by the Stanford University Natural Language Processing Group to determine the proportion of all syntactic tokens with each part of speech for the titles of works created in each decade.⁴² For tokens such as punctuation marks and numerals the tagger effectively identifies each appearance. For other tokens such as nouns and verbs that can be syntactically ambiguous the tagger looks to resolve that ambiguity by considering features relating to the token itself, for instance word endings, and to the context in which the word appears. The tagger assigns weights to each feature to determine how likely a token is to be of each part of speech. The values of the weights are determined by calibrating the tagger against a pre-tagged text, and when run against a new text it assigns the most likely part of speech to each token.

My topic modelling and parts-of-speech analysis involved looking at the titles in each decade as a bag of words, with no other consideration given to the context in which those words are used. To interpret properties of whole titles rather than their constituent parts in isolation I have drawn on some count-based statistics similar to those used in my first case study.

Looking at these measures helps in disentangling the various uses to which numerals and 'untitled' were put and at how artists have engaged with the nominative and semantic functions of the title. I read the prominence of numerals and of the term 'number' in topic 2 as indicating, at least in part, a growing interest during the middle decades of the twentieth century in minimising the role of the title in contributing to the meanings given to the work it names. I also commented my interpretation was provisional and incomplete as the context of use was not being taken into consideration, with such a use of numerals being most likely in short titles such as 'construction 1' and 'number 2' rather than in longer titles. I also read the use of the

⁴² The Stanford University Natural Language Processing Group Part of Speech Tagger can be found at <https://nlp.stanford.edu/software/tagger.shtml>.

simple title 'untitled' as another means of minimising the semantic role of the title. It is these 'non-referential' uses of titles that John Welchman has examined in his history of minimal titling.⁴³ However, if we look at how the word length of titles has changed and at the words used in them, we can see these uses of numerals and of 'untitled' as part of a much broader trend of 'minimal' titling extending back earlier into the century and forward to the 2000s. The semantic role of titles was also lessened through reducing the number of words in them, for instance to simple descriptions of the subject matter or to generic labels. Titles such as 'landscape' and 'mother and child' were in common use in the early decades of the twentieth century, with short titles such as 'painting' and 'abstract composition' frequent in the 1940s and 1950s.

A consolidated measure of these trends is that of the prevalence of short titles of three words or fewer in length, split into 'untitled' alone, short titles including a numeral, and other type of short title. These trends are shown in Chart 5.5 and, as can be seen, minimal titling has always been a major component of titling practice. The proportion of all titles of length three words or fewer rose persistently and markedly from 61% in the 1900s to 75% in the 1950s. It has remained at around 70% since then. Within that overall trend, the use of short descriptive titles or simple generic labels peaked in the 1920s. Since then, that kind of use has fallen back to around 50% of all titles, as minimal titling has involved more use of short titles with numerals and of 'untitled'. Both have accounted for around 11% of all titles since the 1980s.

⁴³ Welchman, (1997).

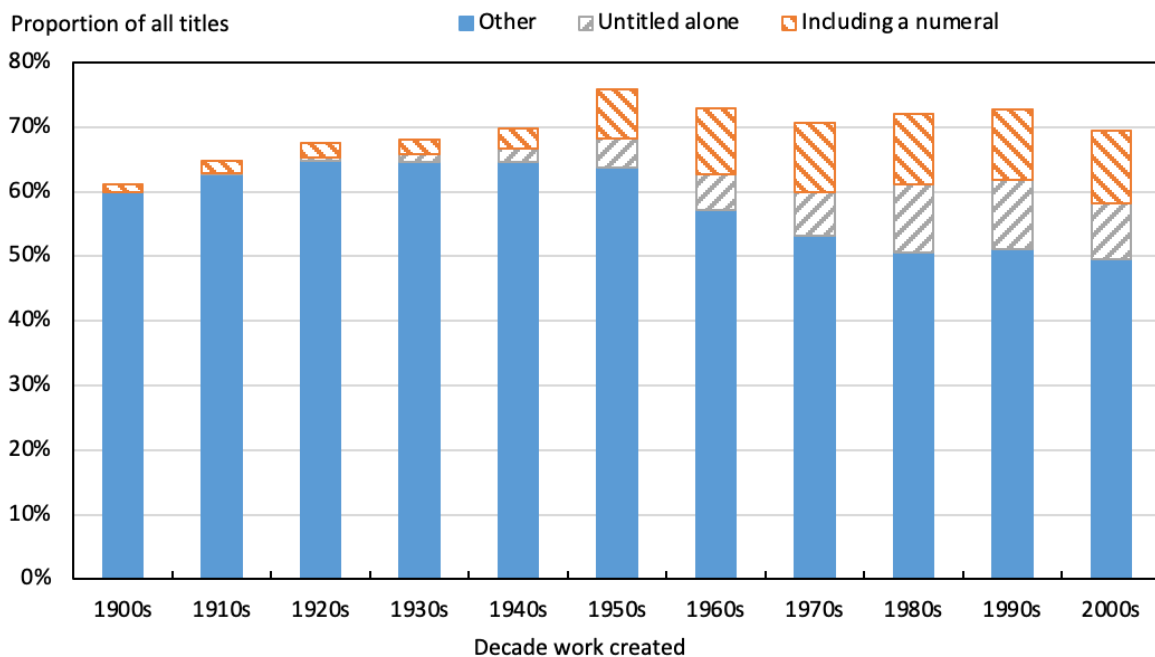


Chart 5.5. Titles of length one, two or three words, 'untitled' alone, including a numeral, and other, proportion of all titles, 1900s to 2000s.

In my reading of topic 3 I suggested that all uses of 'untitled' engage with the nominative function of the title, but not always with the semantic role. When used with other text in the title rather than on its own, 'untitled' most likely functioned to introduce a comment on the work without naming it. One example is the artist Dan Flavin, who often used bracketed text in titles such as *untitled (to Don Judd, colorist)* as personal dedications. The growth in prominence of topic 3 in my topic model indicated that artists were using 'untitled' more commonly, but did not allow the two uses of the term that engaged with the semantic and nominative functions of the title to be distinguished. The more complete picture given by count-based statistics can be read as confirming and showing the relative strength of these trends. Looking first at all uses of 'untitled', Chart 5.6 shows that this became significant in the 1930s and 1940s, and has continued to grow since then, with around 14% of titles in the 1990s including the term 'untitled', whether on its own or as part of a longer title. It also shows there has been a shift in recent decades towards the use of 'untitled' as part of longer titles, which has grown whilst the use of 'untitled' alone has remained largely static. Artists have retained an interest in communicating meaning through the titles they give to their works and have increasingly done that indirectly through the text accompanying 'untitled' rather than through the sort of direct statement that was the predominant form of titling earlier in the twentieth century.

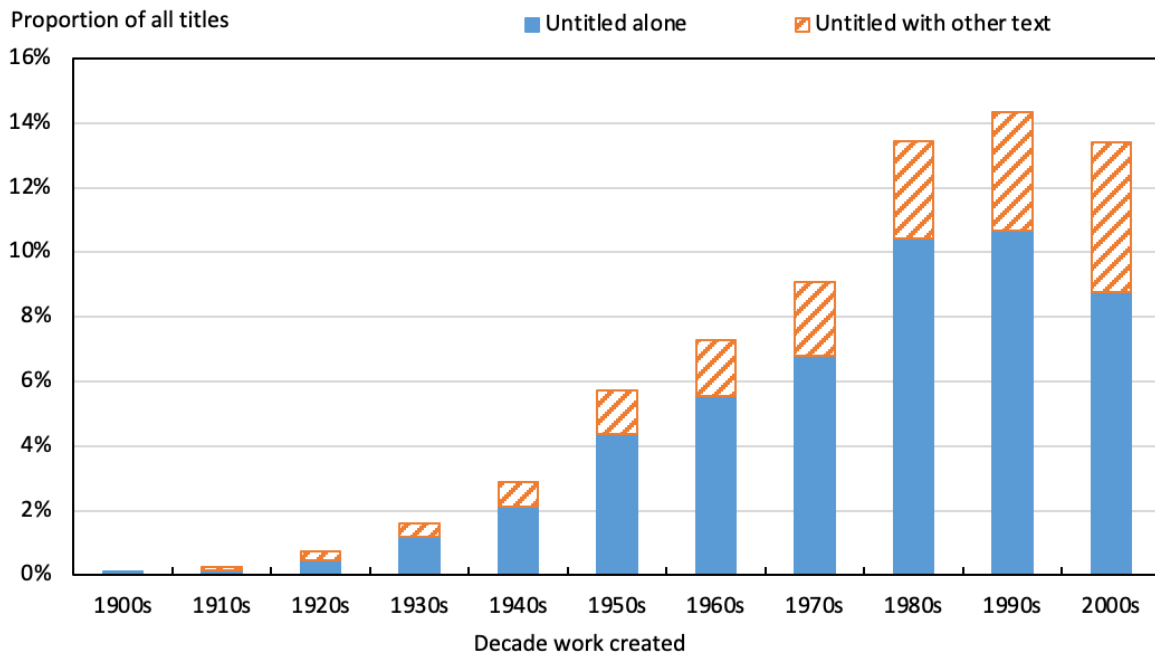


Chart 5.6. Use of 'untitled', on its own and as part of a longer title, proportion of all titles, 1900s to 2000s.

Titling has also become more polarised in recent decades. At one extreme, minimal titling has remained a major part of titling practice. At the other, the use of long titles and the syntactic complexity of titles have both been growing. After having declined in the first half of the twentieth century, use of long titles of ten words or more then picked up, growing from under 1.0% of all titles in the 1950s to 4.0% in the 2000s, as shown in Chart 5.7. The parts-of-speech model shows that in recent decades there has been more variety in the parts of speech used in titles, including increased use of all kinds of punctuation mark. A succinct measure of the syntactic complexity of titles in this sense is the entropy of the parts-of-speech distribution, and, as can be seen from Chart 5.7, this has been rising slowly and steadily. In my first case study I noted that, at least when first used, short entries in the catalogues of the Paris Salon would have stood out from those around them and so functioned 'seductively', attracting the attention of the reader. The increased use of long titles and of devices such as exclamation marks and brackets will have had a similar effect in recent decades.

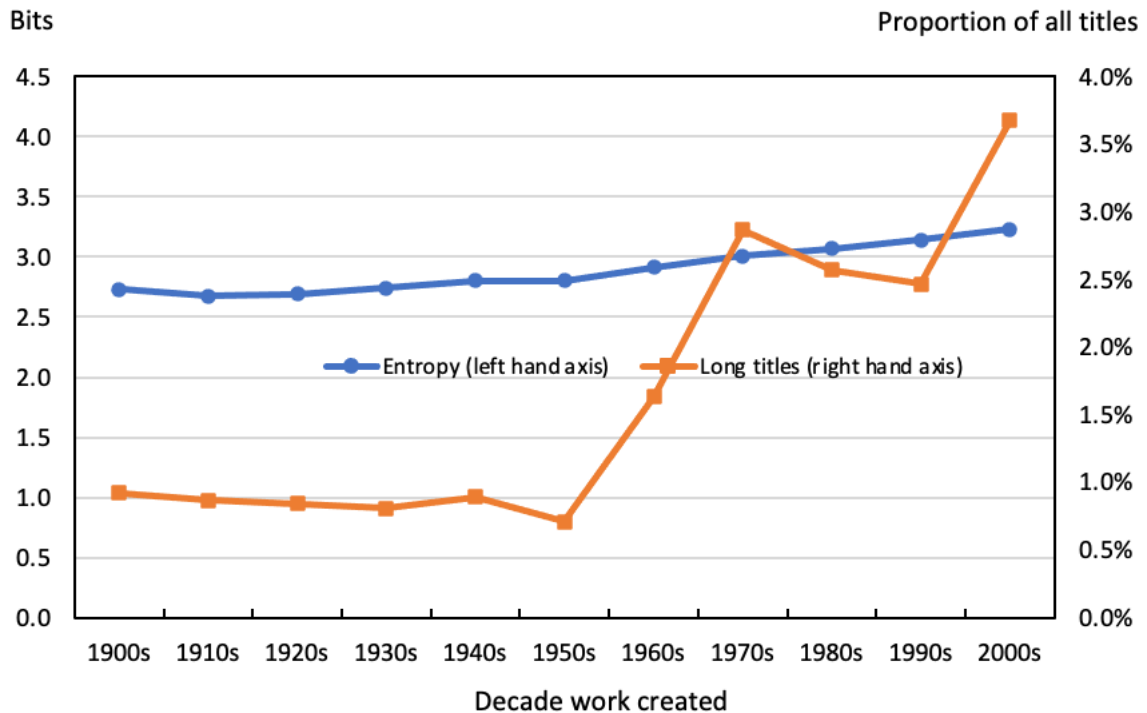


Chart 5.7 Entropy of the parts-of-speech distribution, bits, and titles of length ten words or more, proportion of all titles, 1900s to 2000s.

The parts-of-speech statistics also provide other perspectives on the ways titles were used. In particular, the use of different kinds of word and other syntactic components such as punctuation marks can be interpreted as indications of epistemic perspective. In the first half of the twentieth century the predominant parts of speech were nouns, prepositions, determiners, and adjectives. The sorts of titles they were used in were typically declarative or descriptive such as *Interior with Violin*, *The Seine near Paris* or *Composition in Red* and so expressive of a static, objective, or third-person epistemic perspective. This was the dominant perspective on the kind of knowledge art can or should produce expressed through titles in the first half of the twentieth century. Since the 1950s that stance has been changing to be one that is more dynamic, subjective, questioning, and opinionising. There has been a modest but steady move away from declarative and descriptive titles to the subjective, interrogative, exclamatory or imperative title. This has happened through an increase in the use of personal pronouns, of end-sentence punctuation such as the exclamation mark, the question mark, and the ellipsis, of modals such as ‘can’ or ‘should’ and ‘wh-words’ such as ‘who’, ‘what’, ‘where’ and ‘how’. Examples of this kind of title include *Listen and be Quiet!* and *May I Help You?*. The growing use of text in brackets, for instance following ‘untitled’, is another way this has happened. The use of verbs and adverbs in titles has also grown. Through being used in titles

such as *Succeeding the Past*, I would read this trend as showing an increased interest in process, agency, and change. Often verbs and other parts of speech have gone together in titles such as *How to Make a Refugee* and *We Could be Looking for the Same Things*.

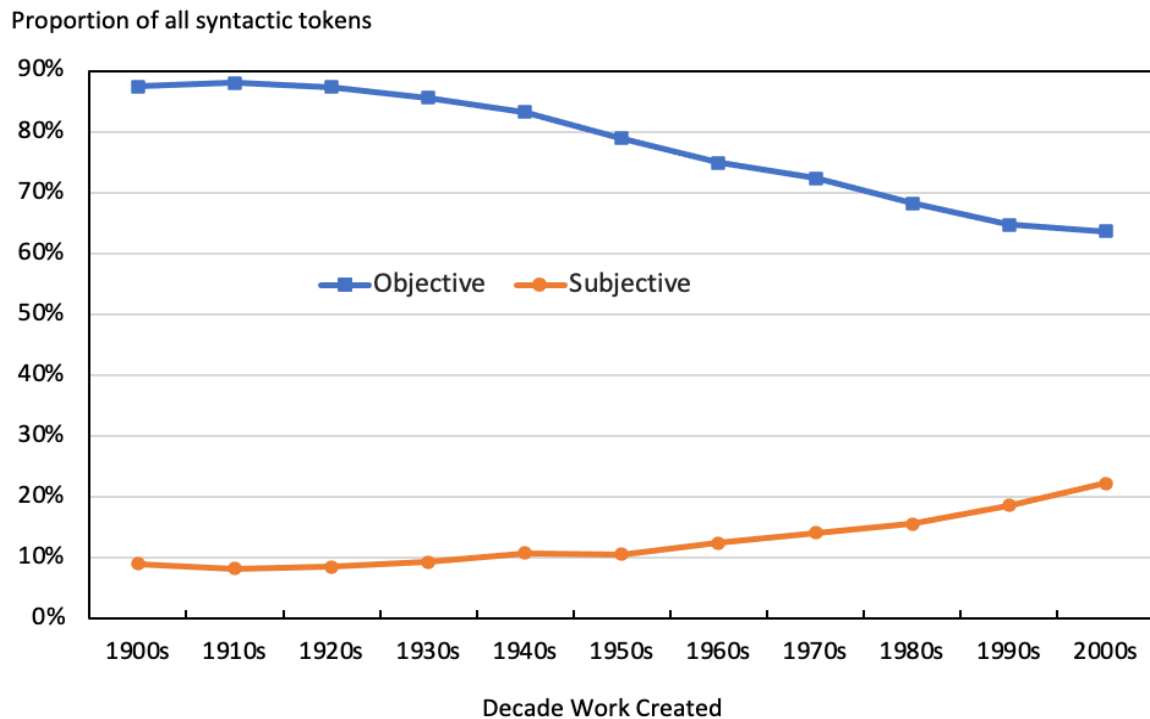


Chart 5.8. Subjective and objective parts of speech, proportion of all syntactic tokens, 1900s to 2000s.

The overall trends in the use of these different parts of speech can be seen in Chart 5.8. For ease of presentation, I have labelled nouns, prepositions, determiners, and adjectives taken together, but excluding all instances of ‘untitled’, as ‘objective’. I have labelled brackets, personal pronouns, end-sentence punctuation, modals, verbs, adverbs, and all instances of ‘untitled’ followed by additional text as ‘subjective’. The remaining syntactic tokens are predominantly ‘untitled’ used on its own and numerals, which I have excluded as I would not read them as strongly suggesting one epistemic stance or another. However, this labelling needs to be interpreted with care. Nouns, verbs, and adjectives can feature in titles presenting a subjective viewpoint as can verbs or adverbs in declarative titles. As with many of the other measures I look at in this case study, it is the relative levels of use that can be read as suggestive of the changing epistemic stance. As can be seen ‘objective’ parts of speech have fallen from nearly 90% of all syntactic tokens in the 1900s and 1910s to fewer than 70% in the 2000s. ‘Subjective’ parts of speech have risen from around 10% of all syntactic tokens in the 1940s to 22% in the 2000s.

5.4 The Circulation of Ideas within Modern and Contemporary Art

Artists of different nationalities

In the previous section I set out a reading in which the words and the syntactic structure of titles functioned as signals of artistic interests and epistemic positions. The framework I developed is not limited to looking at titles in aggregate. It can be applied to the titles in my dataset grouped in other ways, and so allows a comparative reading of how titles were used across those groups. One way is to look geographically. On a strict reading, this would give a comparison of the collections of modern and contemporary art in each of the countries included in my dataset. However, my aim in this thesis is not museological but art historical and so I have taken a different and more suggestive and speculative interpretive approach. Most of the institutions in my dataset have a bias towards native or naturalised artists or are specialised collections of the art of a particular region, and so if we split the museums in my dataset geographically, we can get a perspective on the circulation of ideas between artists of different nationalities. To do this I grouped the museums up by country or, where the number of titles was too low for a robust analysis, combined countries or institutions into regions based on geographical and cultural proximity. One of the regions was Latin America, which included the Museum of the Americas, the Daros collection, and the collections of Latin American Art in the Los Angeles County Museum and the Houston Museum of Fine Arts, and the other was Iberia.⁴⁴ I excluded the Guggenheim from my analysis, as it includes work from the New York, Venice, and Bilbao collections. Altogether this resulted in nine countries and two regions.

There are very few titles from 1900s and 1910s in the Latin American, Korean, and Japanese collections in my dataset, and very few domestic artists in several others including some of the museums in the United States. For these reasons, I excluded the 1900s and 1910s from my analysis. The total number of titles in the collections of each country or region used in my analysis is given in Table 5.4.

France	Germany	Iberia	Italy	Japan	Korea	Latin America	Sweden	The Netherlands	United Kingdom	United States
10,130	4,492	3,172	1,463	4,985	3,108	3,483	4,689	2,511	4,933	8,432

Table 5.4. Total number of titles for each country or region.

⁴⁴ For a discussion of the emergence of Latin America as a cultural category in the visual arts and as a term used by artists to self-identify see Michele Greet, *Transatlantic Encounters: Latin American Artists in Paris between the Wars*, (New Haven: Yale University Press, 2018), pp. 56 - 79.

To model the parts of speech and to produce count-based statistics for titles split by decade and by geography I followed the same approach as when looking at all titles. However, with topic modelling there is no guarantee that identically the same topics will be produced when modelling a different collection of texts. To ensure this, and so allow my interpretations of topics 1 to 3 to be carried over I used the MALLET application in 'inference' mode, in which it took those topics and matched them against the titles used in each country or region and in each decade. For some countries and for some decades the number of titles is low, and the results of the topic modelling and of the analysis of parts of speech exhibited significant variation from one decade to the next. To bring out the trends and to reduce the decade-to-decade volatility in the results I found it was best to combine each decade with the next in what is known as a 'moving average'. I will look first at the overall trends and then consider what my analysis suggests about differences between artists of different nationalities.

In my topic model there were three prominent topics. Topic 1, which I read in terms of an engagement with figuration, was dominant in the early twentieth century. In the middle decades topic 2 was the most prominent, whose most important words I interpreted as signals of interests in the abstract/formal dimensions to art, along with those associated with the use of numerals. Topic 3, which showed the increase in the use of 'untitled' alongside a growing interest in seriality and a range of other concerns, was the dominant topic in the decades from the 1970s onwards. The same inter-woven pattern appears when looking at national or regional collections, suggesting that artists of different nationalities have followed similar paths, as can be seen from Charts 5.9, 5.10 and 5.11.

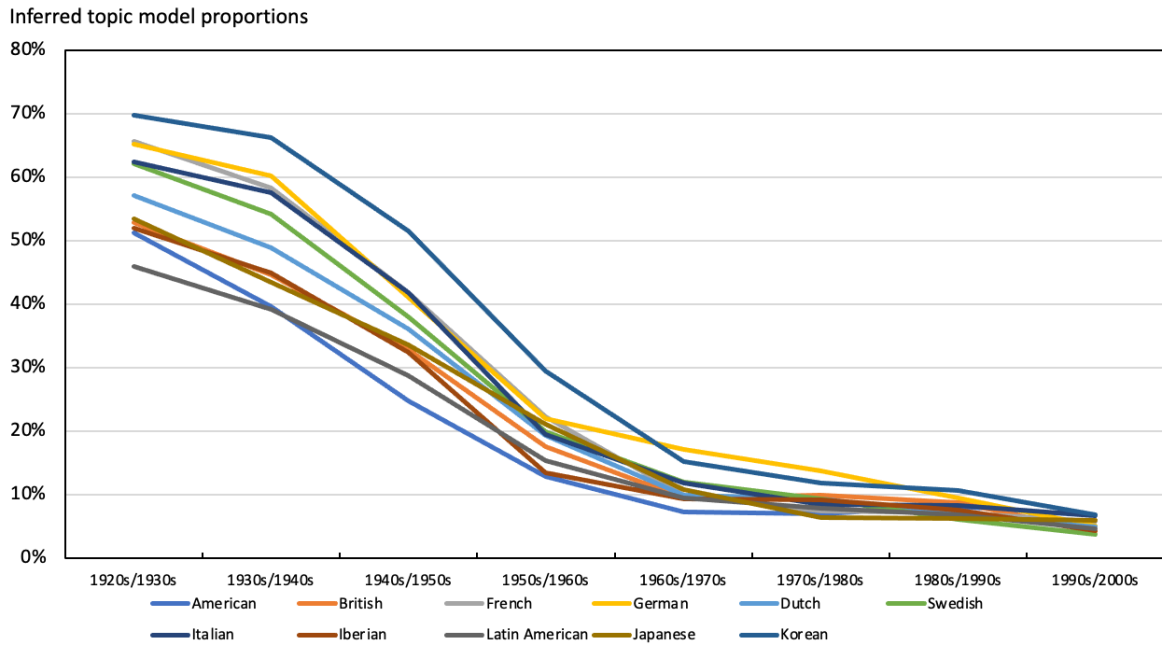


Chart 5.9. Topic 1 proportion in the inferred model for national or regional collections, two-decade moving average, 1920s/1930s to 1990s/2000s.

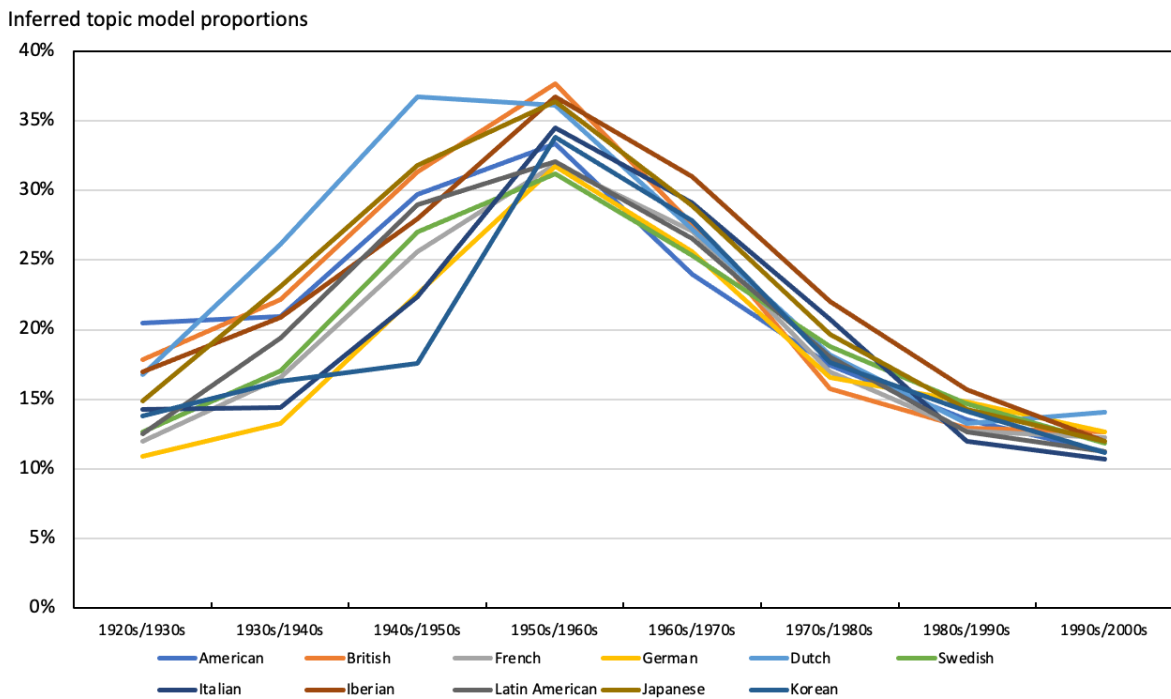


Chart 5.10. Topic 2 proportions in the inferred model for national or regional collections, two-decade moving average, 1920s/1930s to 1990s/2000s.

Inferred topic model proportions

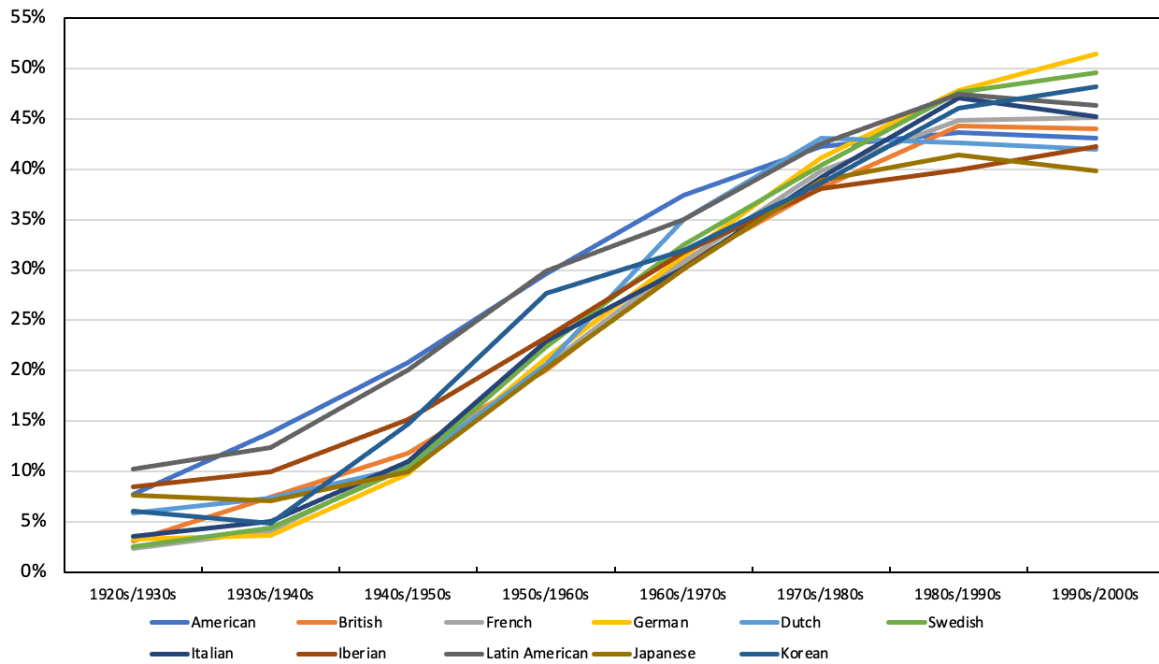


Chart 5.11. Topic 3 proportions in the inferred topic model for national or regional collections, two-decade moving average, 1920s/1930s to 1990s/2000s.

When looking at the parts of speech used in titles and at various count-based statistics I developed a reading in which I looked at how artists had engaged with the functions of the title. I identified a persistent trend through the twentieth century of ‘minimal titling’ looking to downplay the semantic importance of the title. I read this from the increasing use of short titles of three words or fewer in length, including those with a numeral and 'untitled' alone. It is not feasible to present graphically each of these three types of short title for the national or regional collections in a way in which all can be compared visually, but the overall trends can be presented. I do this in Chart 5.12. As can be seen, it indicates that artists of each nationality have followed similar paths with minimal titling rising somewhat in the first half of the twentieth century before peaking in the 1950s or 1960s and remaining largely stable since then.

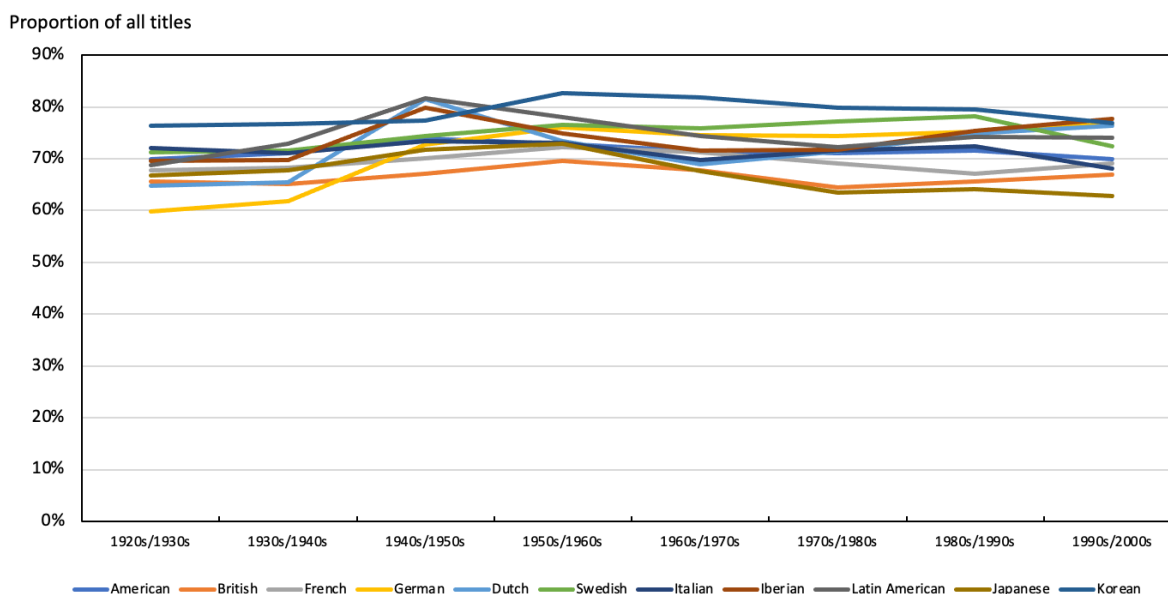


Chart 5.12. Short titles by country or regional collection, proportion of all titles, two-decade moving average, 1920s/1930s to 1990s/2000s.

I also discussed how artists engaged with the nominative function of the title through the use of 'untitled', either in combination with minimal titling by using it on its own or looking to continue to say something about the work by including it in a title with additional text. The overall levels of use of 'untitled' is presented in Chart 5.13. This is the measure on which the differences between national and regional collections are the most pronounced. Although use has risen with all, there is a wide spread relative to the average, and my analysis suggests German artists are far more likely to present works as untitled than artists of other nationalities.⁴⁵

⁴⁵ The much higher use of 'untitled' by German artists than those of other nationalities is one reason why German artists feature prominently in the group of artists I look at in my second case study.

Proportion of all titles

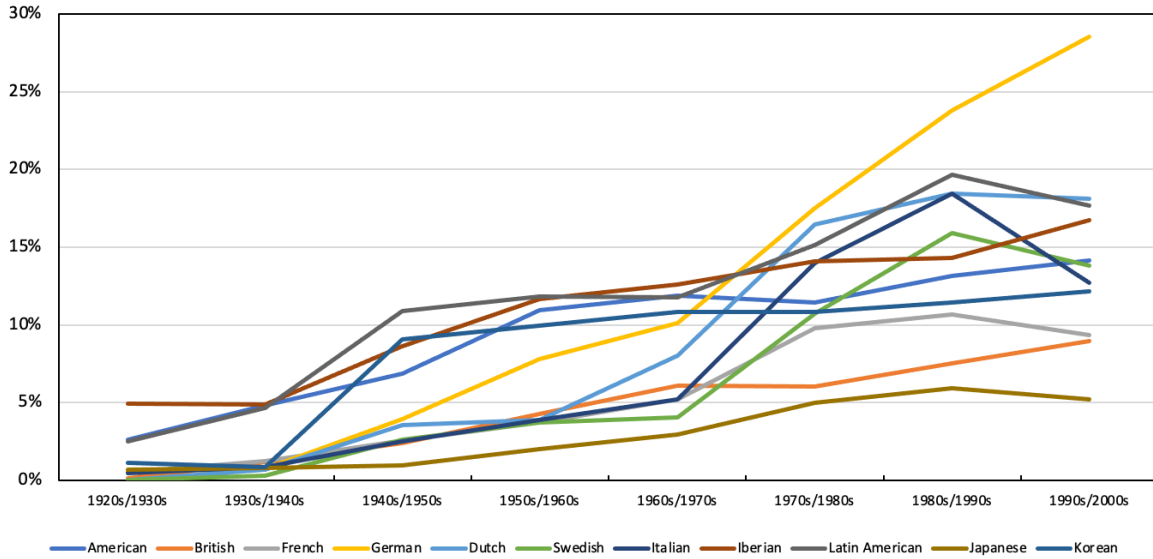


Chart 5.13. Titles including ‘untitled’ for country or regional collections, proportion of all titles, two-decade moving average, 1920s/1930s to 1990s/2000s.

In the previous section I interpreted titles in the early to middle decades of the twentieth century as predominantly expressing a static and objective or third-person epistemic stance. I read this from the preponderance of titles that were descriptive or declarative, as indicated by the very high levels of use of nouns, prepositions, determiners, and adjectives (‘untitled’ excluded). I also read the growing use in recent decades of parts of speech such as question marks, conditionals and wh-words titles as expressions of an increasingly subjective, imperative, and questioning stance, and of growing interests in process and change. We can see the same trends with artists of different nationalities, as indicated by Charts 5.14 and 5.15, which present the ‘objective’ and ‘subjective’ measures for the national or regional collections.

Proportion of all syntactic tokens

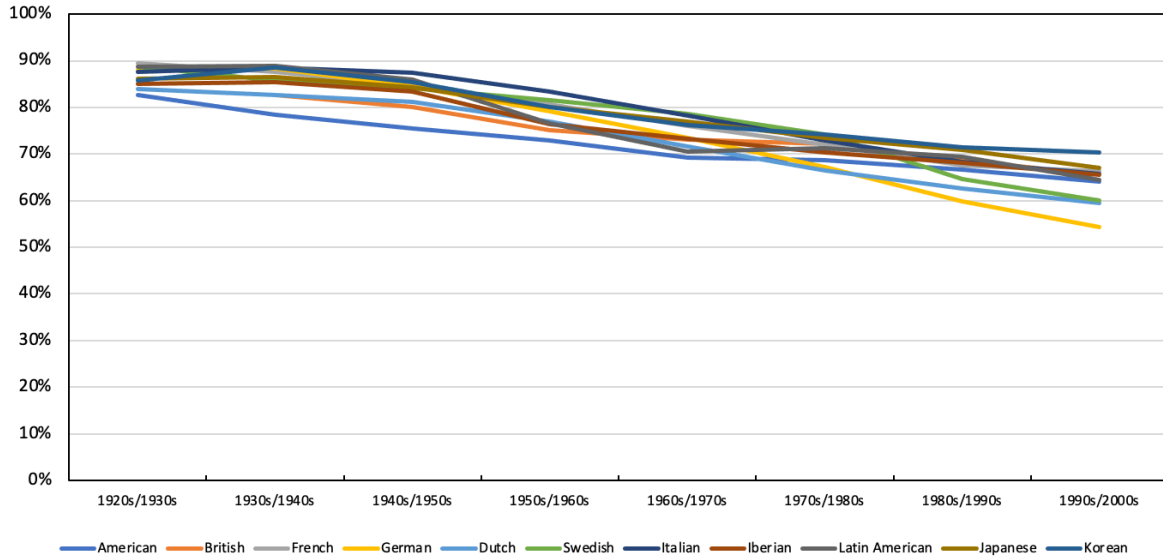


Chart 5.14. Objective parts of speech, proportion of all syntactic tokens for each country or regional collection, two-decade moving average, 1920s/1930s to 1990s/2000s.

Proportion of all syntactic tokens

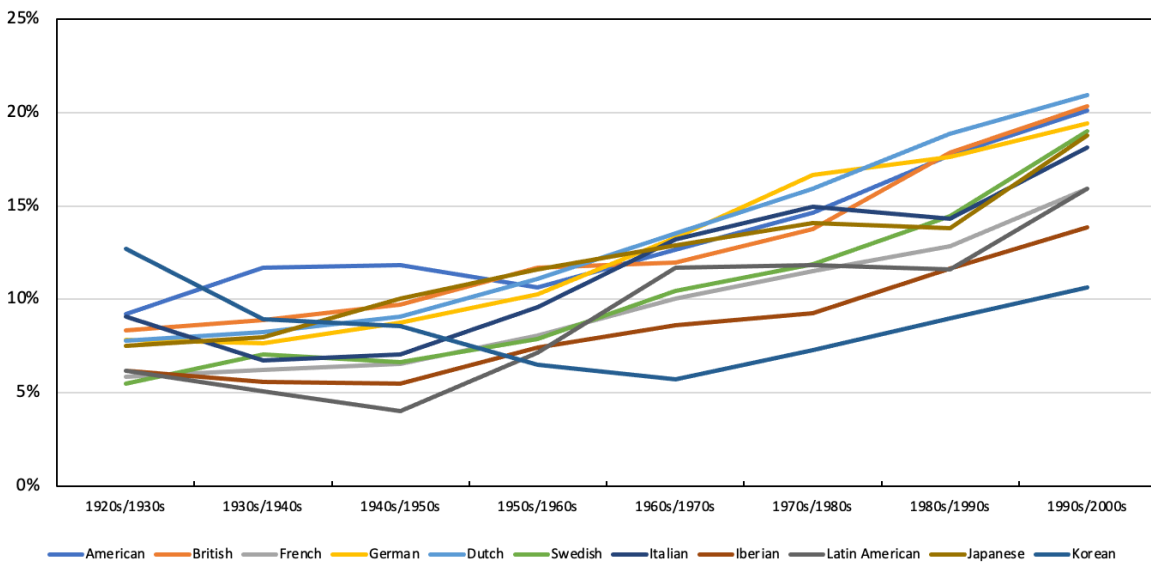


Chart 5.15. Subjective parts of speech, proportion of all syntactic tokens for each country or regional collection, two-decade moving average, 1920s/1930s to 1990s/2000s.

The readings I have presented so far in this section suggest there was a close and parallel alignment between artists of different nationalities throughout the twentieth century in the ways in which I have read titles as being used as signals of ideas and interests. To complete my reading, I will now consider the question of any differences between nationalities. If we look

first at the inferred topics as shown in Charts 5.9 to 5.12, we can see two trends when comparing national or regional collections. Firstly, in recent decades there has been more variability in the orderings in each decade, with no particular collection being consistently amongst the strongest or weakest, compared to the early and middle decades of the twentieth century, when the rankings by decade were more stable. There has also been less dispersion in the distribution of inferred topic model proportions in each decade. A standard measure of dispersion is the 'coefficient of variation', or the ratio of the standard deviation to the average of a distribution. With the inferred proportions for topics 1 and 3, the coefficients of variation have consistently fallen, and with topic 2, the coefficient is lower for recent decades compared with earlier decades in which the inferred proportion is at the same level. I would read these trends as suggesting greater homogeneity on these measures between artists of different nationalities in recent decades than earlier in the twentieth century.

With the measures I used of minimal titling, of use of 'untitled' and of titles signalling an objective or subjective epistemic stance, the rankings across decades have been more consistent than with topic 1, 2 and 3. It indicates that minimal titling was most common with Iberian, Korean, Swedish, and, since the 1950s, German artists. It was used least by British, French, and Japanese artists. Along with German artists, presenting works as untitled was most common with Dutch, Iberian, and Latin American artists, and least common with those of British and Japanese nationality. The objective stance was strongest with Italian, Japanese and Korean artists, and weakest with American, Dutch and, in recent decades, German artists. The subjective stance was strongest with American and Dutch artists and weakest with Iberian, Korean and Latin American artists.

Whilst these patterns of increasing homogeneity with topics 1, 2 and 3 and of consistent high or low levels with other uses of titles are suggestive, it is beyond the scope of this thesis to consider if there is anything of art historical interest in these comparisons. Looking at metadata on its own is insufficient to do this, which would require a much more detailed examination of the histories of artists of different nationalities.

Before summarising the reading I have given of artists of different nationalities, I would like to consider an alternative explanation of the increasing homogeneity between artists I read from my inferred topic model, which is that it is the collections in my dataset that have become more homogeneous rather than the artists of each nationality. If that were the case, you would expect there to be a greater commonality of artists across the works in the collections for recent decades than for earlier decades. To investigate this, I looked, decade by decade, at

the degree of overlap of the artists in each of the collections. To allow for comparisons between decades I used the same number of titles in each, randomising the order of titles for decades with more than this number of titles in my dataset. Having compiled a list of the artists represented in any of the collections in each decade I then used a 'vector space' representation of each country or region identifying whether each artist was in that country's collection. Using the cosine measure of the similarity of each pair of collections in this representation, I calculated an overall average for each decade as a summary measure of similarity, and looked at the standard deviation of those cosine similarities.⁴⁶ Both of these measures changed very little decade by decade, which indicates that the collections themselves have not become more homogeneous.

To summarise, at the scale given by my topic modelling and parts-of-speech analysis, and from the perspective given by my interpretation of them, what is striking when the analyses are repeated by country or regional collections is that from the 1920s developments have been global and largely synchronous. The circulation of the ideas and interests expressed through titles has also been characterised by an increasing homogeneity on several measures. Although the differences are typically small there are also some cases where my analysis suggests that artists of one or more nationality were persistently stronger or weaker on one measure than the rest. As with my other case studies, we can see how the methods of digital art history expose questions which might be addressed using conventional art historical methods.

Male and Female Artists

As with the circulation of ideas between artists of different nationalities we can apply my analytical framework to give a perspective on some of the ways female and male artists used titles. To do this I divided the titles in my dataset by decade and by gender, and repeated my analyses, using the topic model inferentially and other methods directly on the data. As I have already discussed, women are substantially under-represented in the collections of the modern and contemporary art museums included in my dataset. Although there are sufficient numbers of titles for my analysis, the results for female artists are more volatile than for male artists, as can be seen for example in Chart 5.18. The total number of titles for female and male artists in each decade is given in Table 5.5.

⁴⁶ For an introduction to vector space models and the use of cosine similarity in literary history see Andrew Piper, *Enumerations: Data and Literary Study*, (Chicago: Chicago University Press, 2018), pp. 13 - 17.

	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	Total
Female artists	85	238	395	434	329	513	820	805	934	1,155	1,073	6,781
Male artists	2,261	3,742	4,652	5,066	3,903	5,375	7,856	5,186	5,420	3,577	2,874	49,912

Table 5.5. Titles by female and male artists, 1900s to 2000s.

Inferred topic model proportions

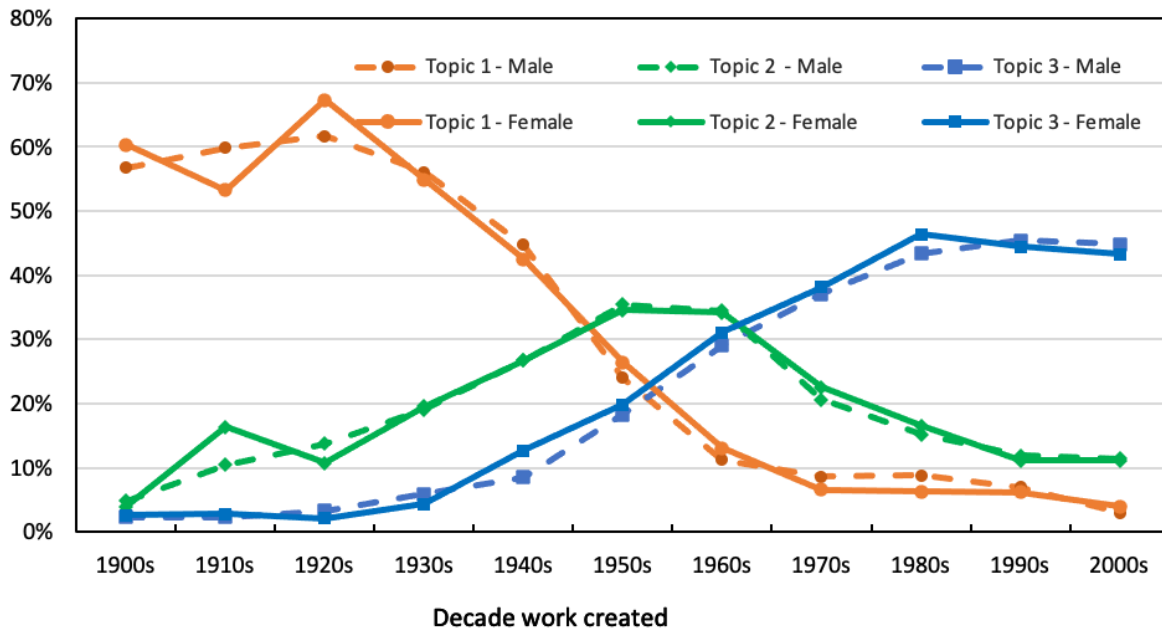


Chart 5.16. Topic 1, topic 2 and topic 3, proportions in the inferred topic model, female and male artists, 1900s to 2000s.

I will first compare male and female artists on the proportions of topics 1, 2 and 3 in the inferred topic model. This is shown in Chart 5.16, where the same inter-woven cyclic pattern of change can be seen with both genders. Titular signals of interests in relation to figuration are dominant in the early twentieth century, before being displaced by emerging interests in abstraction and the formal, which peak in the 1950s and 1960s, and are then themselves displaced by the heterogeneous set of interests indicated by topic 3.

Proportion of all titles with assigned gender

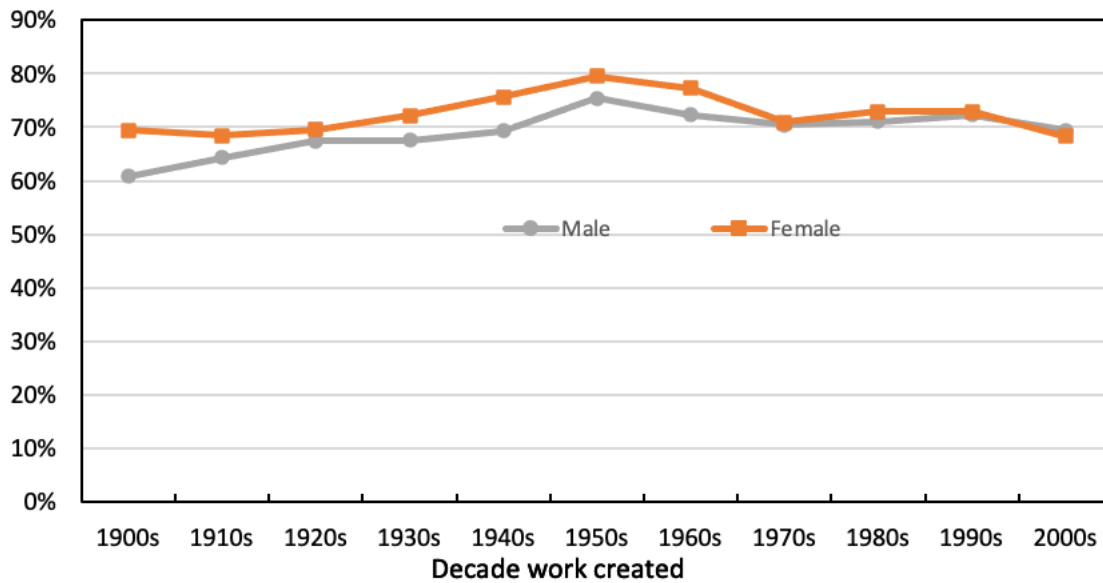


Chart 5.17. Short titles of length three words or fewer, proportion of all titles with assigned gender, female and male artists, 1900s to 2000s.

I will now look at how female and male artists engaged with the semantic and nominative functions of the title. Through the twentieth century artists looked to minimise the semantic role through the use of short titles of length three words or fewer, including numerals and 'untitled' alone. The similar paths followed by male and female artists in their use of short titles can be seen from Chart 5.17. Since the 1940s male and female artists have also engaged with the nominative function of the title, diminishing it through the use of 'untitled' to a similar extent, as shown in Chart 5.18.

Proportion of all titles with assigned gender

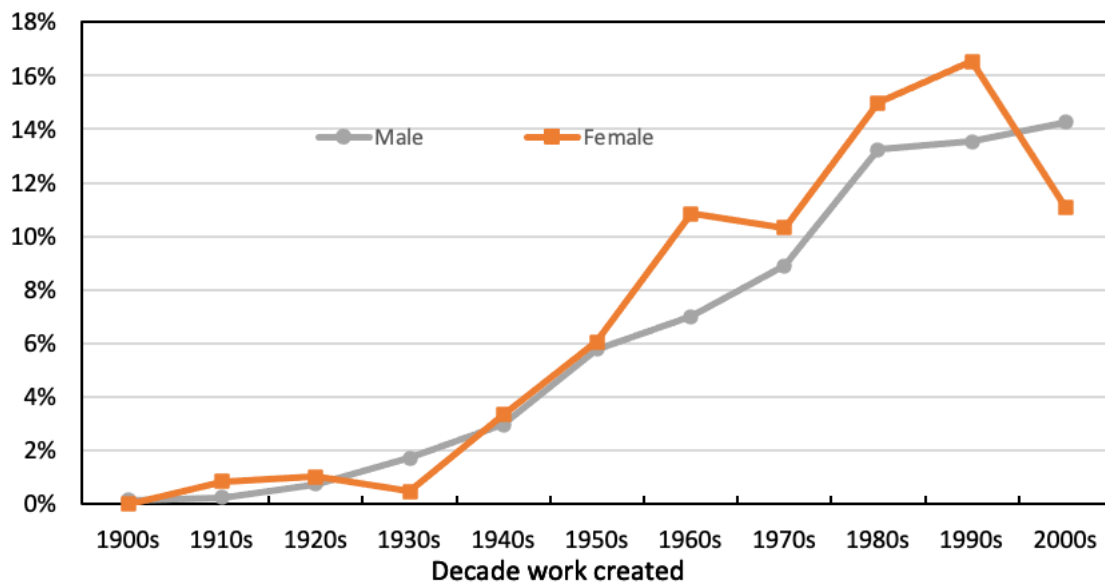


Chart 5.18. Titles including 'untitled', proportion of all titles with assigned gender, female and male artists, 1900s to 2000s.

The shifting epistemic perspective from a predominantly static, objective position to one that was increasingly dynamic, subjective, and questioning and which I read, respectively, from the decline in the use of certain parts of speech and the increase in the use of others along with long titles, is something that can be seen with both male and female artists. As can be seen from Chart 5.19, the measures I have read in terms of the epistemic perspectives signalled through the titles of works followed close and parallel paths with female and male artists throughout the decades from the 1900s to the 2000s.

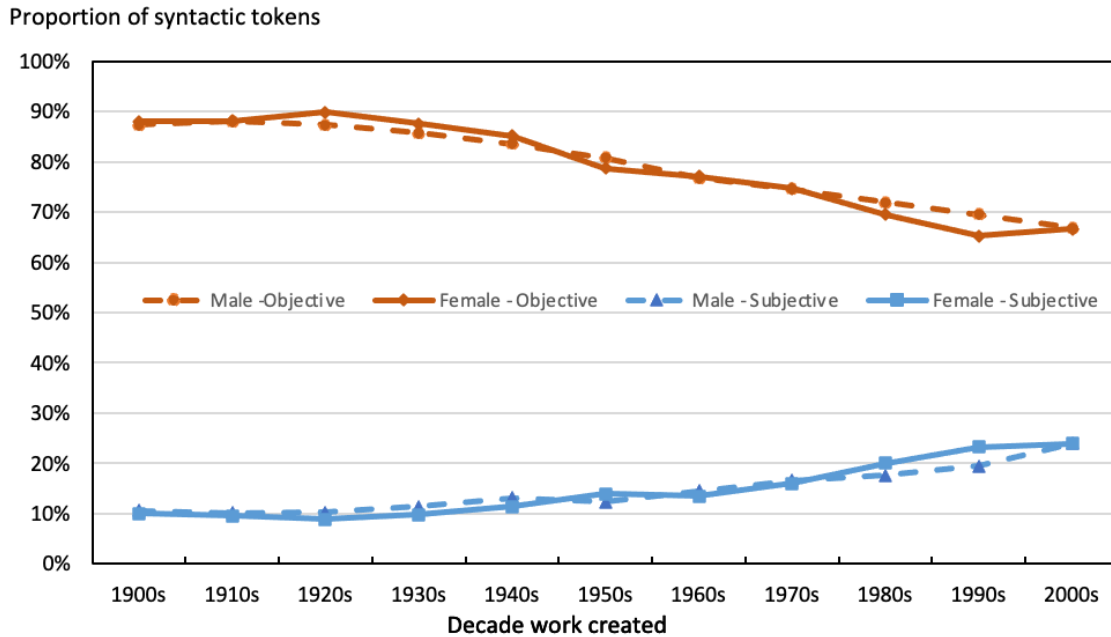


Chart 5.19. Objective and subjective parts of speech, proportion of all syntactic tokens, female and male artists, 1900s to 2000s.

In summary, on all the measure I have looked at of the ways the language in titles was used by artists, female and male artists followed very similar paths throughout the period from the 1900s to the 2000s.

5.5 Summary and Conclusions

Overview

In this chapter I have set out my third case study into the application of digital resources and statistical methods in art history. I will first give an overview what I have done before considering the answers it gives to the broad questions which have guided my work. The dataset I compiled for this case study was the metadata of the artist's name, the work's title and the year of creation for over 59,000 entries downloaded from the online collections of 35 modern and contemporary art museums. In contrast to my first two case studies, I have brought a range of statistical techniques to bear on my data using them individually and in combination to develop my readings. In this instance the predictive text-mining methods of topic modelling and parts-of-speech tagging, along with the descriptive methods of correspondence analysis and word counting. I have also used the information theoretic concept of entropy as a measure of the diversity and complexity of titles.

Correspondence analysis showed some strong trends in the use of titles through the twentieth century but has its limitations from an art historical perspective, and so I turned to other text-mining methods. In my reading of the results of the topic model we can see an inter-woven cyclic pattern of change in which the use of titles to express certain interests increases in importance, becomes dominant and is then displaced. Titular signals in relation to figuration are displaced primarily by those relating to abstraction and the formal dimension to painting, which in turn give way to the more heterogeneous set of interests expressed through the words in topic 3. What cut across the expression of the two types of interest dominant in the first half of the twentieth century was an epistemic stance in which titles expressed an objective and stable perspective. In the second half of the twentieth century and into the twenty-first, the kinds of interest I associated with topic 3 were associated with a more subjective and questioning epistemic position.

Titles have been used by artists to engage with the three basic functions of the title. Minimal titling, which looked to reduce the role of the title in the meanings given to the works they name, was an important element of titling throughout the twentieth century. Presenting works as untitled is a way of engaging with the nominative function of the title. It first became significant in the 1930s and 1940s, and has grown strongly since then. In recent decades there have been increases in the use of very long titles, and in their syntactic complexity. Such titles stand out from those around them, and so 'seduce' the reader.

I also use the interpretive framework I have developed to compare artists of different nationalities, and male and female artists. What this suggests is that no particular nationality was a leader or follower, in the sense that interests or ideas expressed through titles begin to grow or fall and get significantly stronger or weaker with those artists much earlier or much later than with the rest. In terms of the uses of titles I have investigated, male and female artists also followed similar paths.

Digital Art History

Online museum collections and the statistical methods I have utilised to model that data have received limited attention in digital art history. The readings I present show their art historical value. As with my other case studies, they also show how the resources and methods of digital art history allow complex inter-relationships to be investigated. In this case, I apply the analytical and interpretive framework I have developed to artists of different nationalities and genders.

This case study sets out a way of thinking about some of the history of modern and contemporary art in terms of ongoing artistic concerns and interests which rise and fall in importance. It is consistent with the canonical history of modern and contemporary art and cuts across the particularities of artist, period or movement which often feature in such accounts.⁴⁷ It is not possible to summarise and adequately reflect the complexity of that history in a single paragraph, but I will suggest some of the ways we might compare them. Whilst some artists in the early twentieth century continued to work within the figurative traditions of the nineteenth century, others in the main avant-garde movements of that time such as Fauvism, Cubism or Der Brücke challenged that tradition through giving different ways of seeing or representing figurative subject matter. Interests in regard to figuration continued to be a concern for many modern artists after the First World War, for instance with the revival of Realism in the 1920s. Artists who combined natural forms to present an alternate and subversive reality were one important strand within Surrealism from the 1920s onwards. The emergence of abstract art as a significant trend within Modernism is often traced back to the 1910s and associated with artists and movements such as Wassily Kandinsky, Piet Mondrian, De Stijl and Suprematism. Abstraction grew in importance in the 1920s and 1930s, and from the 1940s to the 1970s it was a widely practiced form of Modernism, including various strands or schools of expressive and geometric abstraction such as Abstract Expressionism, Tachisme, Art Informel and Concrete Art. The pattern of development of topics 1 and 2 in my model matches that narrative. Since the 1960s art has been characterised by an artistic plurality. Working with a wide range of media artists have engaged with a diversity of styles and themes. Some have looked to create a dialogue with styles or movements that came before them, and here we might see the continued significance of topics 1 and 2 in my model. The kinds of interests I associated with topic 3 and the changing epistemic perspective can also be seen in relation to recent developments. Many artists have engaged with seriality as an element of their practice. Other important themes include those of race, personal identity and politics, and the making of art that is explicitly challenging and questioning of institutional structures and power relations.

⁴⁷ For surveys of modern and contemporary art see Bois et al, (2004), Hopkins, (2018), Smith, (2011), and Peter R. Kalb, *Art Since 1980: Charting the Contemporary*, (London: Lawrence King, 2013).

Another art historian who has looked at the use of titles within modern and contemporary art in terms of long-running themes is John Welchman. My reading extends his account of non-referential titling to include the use of all kinds of short title, and complements his discussion of artistic engagement with the compositional.

My reading of the circulation of interests and concerns between artists of different nationalities engages with the work of other art historians. It complements the growing body of work by scholars who have looked beyond the canonical history to explore the ways artists of many nationalities were not merely peripheral but integral to the history of Modernism.⁴⁸

Studies of women artists in modern and contemporary art may look at the contributions of particular artists, consider how women engaged in particular movements or tendencies, or look at how female artists explored particular issues or themes in their art.⁴⁹ What my framework adds to those accounts is to give a viewpoint from which male and female artists are looked at separately but equally. It enables a reading in which, on all aspects such as the expression of interests in figuration or abstraction through the language used in titles and engagement with the functions of the title, female and male artists followed very similar paths.

Statistical modelling of titles also allows me to answer other questions that cannot be addressed through conventional art historical methods. In particular, through an appeal to the information theoretic concept of entropy I have interpreted my model results in terms of the levels of diversity and complexity in the use of titles and in their syntax.

Methodology

There are several methodological lessons I would draw from the approach I have developed for this case study. It differs from my other case studies in that I have carried out much more processing of the data for inclusion in my dataset. After downloading the metadata I processed it in several ways, including the removal of curatorial additions and the translation of all titles into English. I assigned a binary gender to each artist based upon their name. There are also issues of bias with the data in that, for instance, some artists are very heavily represented and others under-represented, and the titles in the online collection may not be the original title of

⁴⁸ See, for example, Greet, (2019), Kaufmann et al, (2017), Beata Hock and Jonathan Owen, *A Reader in East-Central European Modernism, 1918 - 1956*, (London: Courtauld Books Online, 2019), and Elaine O'Brien, Evelyn Nicodemus, Melissa Chiu, Benjamin Genocchio, Mary K. Coffey, and Bernard Tejada, (eds.), *Modern Art in Africa, Asia and Latin America*, (Oxford: Blackwell Publishing, 2013).

⁴⁹ See, for example, Butler Schwartz, (2010).

the work of art. As I have stressed, the researcher needs to be reflective on the consequence of any biases in their data sources and of the ways in which they have processed that data. In the course of presenting this case study I have mentioned several areas where my modelling and interpretive approach provide ways of ameliorating some of these data issues. The translations and gender assignments I have used were adequate for the kind of reading I have presented, and the topics prominent in one decade only often served to isolate the titles of artist who are heavily represented in that period. The compatibility of my reading with the canonical history of modern and contemporary art provides further comfort that the interpretations of the use of titles I set out in this case study are sound.

My work extends the sorts of use to which statistical modelling techniques have been put in the digital humanities. The readings I have developed show that to get the most from topic modelling researchers need to consider the model as a whole rather than focus on particular isolated and univocal topics. Parts-of-speech analysis is typically used in the digital humanities as a first step to further analysis. For instance, to identify the adjectives used in a text as part of an analysis of the sentiments being expressed or to isolate the proper nouns, which can then be extracted as part of the text preparation in topic modelling.⁵⁰ It may also be used for classificatory purposes with the pattern of use of different parts of speech as a marker utilised to identify the genre or author of a text.⁵¹ However, in the digital humanities it has been used only rarely to provide representations that are of further interpretive significance.⁵² In my own reading I have looked to give art historical interpretations to all of the topic model and to all of the information in the parts-of-speech model.

⁵⁰ See, for example, Wai-Howe Khong, Lay-Ki Soon, Hui-Ngo Goh, Su-Cheng Haw, (2018) 'Leveraging Part-of-Speech Tagging for Sentiment Analysis in Short Texts and Regular Texts', in Rayatore Ichise, Freddy Lecue, Takahiro Kawamura, Dongyan Zhao, Stephen Muggleton, Kouji Kozaki. (eds.) *Semantic Technology. JIST 2018. Lecture Notes in Computer Science*, Vol. 11341, (New York, Springer, 2018) and Jockers, (2012), pp. 118 - 153.

⁵¹ See, for example, E. Stamatatos, G. Kokkinakis, and N. Fakotakis, 'Automatic text categorization in terms of genre and author', *Computational Linguistics*, 26, (2000), pp. 471 - 495.

⁵² For one example see Sarah Allison, Marissa Gemma, Ryan Hueser, Franco Moretti, Amir Tevel, Irena Yamboliev, 'Style at the Scale of the Sentence', Stanford Literary Lab Pamphlets, 5, (2013).

Researchers in the digital humanities often look at one quantitative technique, or at several independently or sequentially where one is seen as improving upon the previous. In my work I have moved beyond these types of approach and have used quantitative techniques both individually and in combination to develop a consolidated reading. The various perspectives provided by correspondence analysis, topic modelling, parts-of-speech analysis and descriptive count-based statistics complement each other and, in the cases of the use of the numerals and 'untitled' are required to develop a more complete interpretation.

6. CONCLUSIONS

6.1 Digital Art History

In this concluding chapter I draw together my three case studies and consider the answers that, taken together, they give to the questions I set out in the Introduction that have guided my work. The first group of questions I posed was around the value of digital resources and statistical methods in art history. I began by asking what kinds of art historical question they allow the scholar to address. Looking at digital resources, in principle there is no limit to the size or the scope of the resources that can be drawn on for a digital art historical study. The processing capability of modern computers means that there are few practical constraints as well. In my first and third case studies the datasets I constructed covered extended periods of time; in the first eighty years, and in the second one hundred and twenty. The dataset for my third case study was drawn from the online collections of thirty-seven institutions in nineteen countries. However, this does not mean, as the art historian Paul B. Jaskot has asserted, that digital art history is ‘almost by definition a field that rests on “big data”’.¹ The small sample of critical writing available online that I drew on in my first case study was sufficient for my examination of the emergence of titling. In my second case study I contrasted the very general conclusions that other scholars have drawn from looking at auction market data in aggregate with the fine-grained understanding of collectors’ preferences that can be achieved through the use of separate datasets and models for a small number of artists.

Statistical methods involve measuring the extent to which something happened or was present. As my case studies show, they allow the art historian to address questions of degree at any of the scales present in their datasets from the micro to the macro. In my first and third case studies I provided art historical interpretations to gradual processes happening over the course of many decades and, with the latter, in many countries. In my first case study I also focused in on a sharp change that happened over a short period of time. Whatever the scale of the analysis, this thesis shows how statistical methods can allow the art historian to draw comparisons between particular ‘subsets’ of their datasets, or groups defined in terms of any combination of the features represented in the data. In my first case study, I used descriptive statistical methods to look at the question of how strongly gendered the display of paintings in

¹ Paul B. Jaskot, ‘Digital Methods and the Historiography of Art’, in Kathryn Brown, (ed.), *The Routledge Companion to Digital Humanities and Art History*, (London: Routledge, 2020), p. 9.

oil in different genres was at the Paris Salon and how that changed over time. Also in that case study, I read the language used in catalogue entries for landscape paintings as signaling different kinds of aesthetic value, and so how they were to be appreciated. The prevalence of those signals indicated how the importance and distinctiveness of those kinds of aesthetic value changed over time across the broad community of artist showing landscapes at the Salon. In my third case study I set out a reading of the levels of use of different parts of speech as indicating the kinds of epistemic perspective suggested by the language used in titles. In the same case study I used the interpretive framework I developed through looking at the metadata in aggregate to compare artists of different nationalities.

In this thesis I also use sophisticated explanatory or predictive techniques to go beyond these kinds of comparisons and look at more complex or dynamical inter-relationships. In my second case study, regression analysis allowed me to develop an understanding of the influence of different characteristics of a painting on the price paid at auction. Comparing, for instance, the average prices paid for paintings of different sizes does not show how size on its own influences collectors. Paintings of one size may share other characteristics that are not present to the same extent in paintings of other sizes, and also affect the price. In my third case study, correspondence analysis allowed me to identify some clusters of words featuring prominently in titles during particular periods. I used topic modelling to build on that account and look at the associations of the same word appearing in different topics, which I read as showing how the artistic interests they signaled were re-inflected.

Digital art history, as I have defined it, is methodological, and my case studies show how those methods can be used by the art historian to address two kinds of question. Firstly, cultural questions relating to artistic or aesthetic value, as in my readings of the use of landscape entries in my first case study and of my topic and parts-of-speech models in my third case study. Secondly, contextual questions where art is considered as an embedded practice, such as in my first case study where I read the language used in catalogue entries as being involved in the construction and institutionalisation of certain cultural categories, and in my second case study where I look at art in relation to the auction market.

In the Introduction I also posed the question of whether some statistical methods are better suited to art historical enquiry than others. My case studies were chosen to allow me to consider this question. In the first I drew on descriptive modelling, in the second on an explanatory technique, and in the third I utilised both descriptive and predictive methods. My

third case study also involved the use of the information theoretic measure of entropy. Scholars working in the digital humanities have debated the merits of these approaches to statistical modelling. My case studies support the kind of inclusive approach to method advocated by Andrew Piper, rather than a focus on one kind of method, as argued for by the literary historian Ted Underwood.² They work together to show that all the three main kinds of statistical modelling can be of considerable interpretive value for the art historian. They all have a place in art historical enquiry and should be seen, as Andrew Piper has characterised it, as parts of the digital art historian's 'toolkit'.³

6.2 Digital Art History and Art History

The next group of general questions I raised in the Introduction consider how digital art history stands in relation to the rest of the discipline. Scholars working in digital art history have debated this issue. For Nuria Rodriguez Ortega, there is 'no essential distinction between digital art history and other art histories – there is simply "doing art history"'.⁴ There is no 'radical rupture with the disciplinary tradition', rather what is happening is a 'process of interpenetration in which some things are retained, some are let go, some are reformulated and some emerge as new questions'.⁵ For Harald Klinke, digital art history is still in its early days and its disciplinary status remains an open question.⁶ He identifies four options: it is an extension to art history, it will become the art history of the future, it is a transdisciplinary field sitting between art history and information science, or, it becomes something completely new that stands on its own as a discipline in its own right.

I will come on to give my opinion on the views expressed by Rodriguez Ortega and Klinke after considering what can be learnt from my case studies. Looking first at the question of what new ways of seeing and new knowledge the use of digital resources and statistical methods can bring to the discipline of art history, in this thesis I do things that cannot be achieved through the use of conventional art historical resources and methods. In all my case

² Piper, (2018), p. 101, Underwood, (2019), pp. 190 -193. Underwood prefers predictive techniques of the kind used for classification, arguing these are the kind of method best suited to the sorts of questions posed in the humanities.

³ Piper, (2018), p. 11.

⁴ Nuria Rodriguez Ortega, 'Digital Art History: The Questions that Need to be Asked', *Visual Resources*, 35, (2019), pp. 6 – 20, p. 7.

⁵ Rodriguez Ortega, (2019), p. 7.

⁶ Harald Klinke, 'The Digital Transformation of Art History', in Kathryn Brown, (ed.), *The Routledge Companion to Digital Humanities and Art History*, (London: Routledge, 2020), pp. 32 - 42.

studies the use of statistical methods and the figures, charts, and tables I present are essential to the interpretations I offer. The datasets I have used provide solid evidential bases for drawing general conclusions, and also for more fine-grained analysis.

In my first case study I was able to confirm Ruth Yeazell's tentative and impressionistic observation on the emergence of titling in the French art world in the late eighteenth century.⁷ Yeazell is restricted to pinpointing a moment in time. In contrast, my methods allowed me to move beyond Yeazell and map out the process of the adoption of titling through the nineteenth century. As Diana Seave Greenwald has noted in her work, the use of catalogue entries in my first case study allowed me to go beyond the examination of extant paintings and look outside of the canon.⁸ I was able, for instance, to look at how the dominant ideology of each political regime was manifested in the honorifics used in the catalogue entries for portrait painting. Relying on extant works would not support such a reading. In the same case study, word counting exposed the disappearance of the term 'nature morte' from the catalogue of the Paris Salon. Treating the metadata for works of modern and contemporary art as a whole, allowed me to develop a synoptic and large-scale reading in my third case study that cuts across and contrasts with conventional narratives of artists, movements, or periods.

The linear narrative nature of conventional art historical writing is not well-suited to consideration of the complex multi-dimensional relationships between factors such as language use, subject matter, genre, gender, nationality, and time that I give in my first and third case studies. In contrast, as I observed in my first case study, art historians that have looked at the subject matter of nineteenth-century French genre painting may not consider gender, whereas those looking at women artists may not compare male and female artists or look by subject matter.

The 'all other things being equal', or 'ceteris paribus', mode of thinking with the interpretation of regression models where changes in one characteristic are considered in isolation from changes in the others is not one that has conventionally been used in art historical explication. My second case study shows the art historical value of this way of thinking. Regression

⁷ Yeazell, (2015), p. 48.

⁸ Greenwald, (2021), p. 23.

modelling was necessary for the development of the kind of disaggregated and quantified understanding of collectors' preferences I present in that case study.

The extended readings I gave of the emergence of titling, of collectors' preferences with contemporary art and of what metadata can tell us about the history of modern and contemporary art stand alone as contributions to art history. In other areas my work may be best considered as complementing and extending existing art historical scholarship, giving a fresh perspective on questions such as the gendering of the Paris Salon in the nineteenth century and the circulation of ideas between modern and contemporary artists of different nationalities. My case studies also engage more directly with the preconceptions or the detail of accounts given by other art historians. My account of the changing status of still-life subjects and the contested meaning of the term 'nature morte' problematises the readings given by other art historians who treat still-life as a well-defined genre, signified in French by the term 'nature morte', and encompassing all types of work in the still-life category. In my third case study, the reading I gave of artistic engagement with the semantic function of the title provides a different account of 'minimal' titling to that given by John Welchman.

My case studies also show how digital methods can expose questions that may be best addressed through the use of conventional art historical methods. For instance, the sharp change in the content of still-life entries at the start of the 1870s suggested that the cataloguing and display of still-life subjects at the Paris Salon were elements of the cultural politics at the time that have not been considered by art historians who have written on the period. To take this further would require examinations of the archives of the Salon Administration, of articles in contemporary cultural press as well as Salon reviews and artists' writings.

Taken on their own my case studies would give support to the view expressed by Rodriguez Ortega. My application of the methods of digital art history in this thesis gives new ways of seeing art historical material, and allows new questions to be answered. However, the *kinds* of question I have answered are those which have traditionally been posed by art historians. None of my case studies represents a 'radical rupture' in the discipline, as, for example, feminist art history did from the 1960s onwards. It would be unwise, though, to generalise from such a small number of examples. As I discuss in the literature review, developments in digital image processing and on techniques to model three dimensional objects may suggest new and transformative ways of thinking about art historical material. The inter-relationship between the visual and the verbal is also an area that has received very limited attention. And

so, I agree with Klinko that it would be premature to make any definitive statements on the status of digital art history.

6.3 Methodology

The final general question I posed in the Introduction was around considering my work from a methodological perspective in the context of the digital humanities and the lessons that can be drawn from it. The interpretations I have developed in my case studies show that to get the most from statistical modelling researchers need to consider all the information their models contain. Thematic interpretations of topic models, for instance, ignore the difference in importance of words within topics and the overlap of words between topics. Both of which feature in the readings I developed in my third case study. Thinking about my topic model and parts-of-speech model in these terms was what prompted me to look at the entropy of the probability distributions that feature in those models.

Scholars working in the digital humanities often draw on one statistical technique, or on several independently or sequentially where one is seen as improving upon the previous. This reflects a methodological position in which different methods are considered as alternative ways of approaching the same question or as techniques that address distinct kinds of question.⁹ In my work I have moved beyond that approach and in my third case study have used a range of statistical techniques both individually and in combination to develop a consolidated reading. The various perspectives they provide complement each other and deliver a more complete and richer interpretation than would be obtained through looking at each in isolation. The same considerations apply to my data sources. Rather than restricting my attention to one kind of data such as a literary corpus, in my first case study the readings I developed drew on the language used by artists in their catalogue entries and by critics in their reviews of the Paris Salon. Text mining allowed me to look at how the same words were being used in those two kinds of textual discourse and at the context of that use.

⁹ This is the position taken by both Piper, (2018), and Underwood, (2019).

6.4 Summary

In summary, this thesis presents an extended exploration of the application to art history of statistical methods drawing on textual or numerical resources in digital format. It presents a kind of art writing in which the numerical and the graphical are integral to the readings being offered. Each of my case studies advances digital art history in some of its less-developed areas. Over and above the individual contribution of each case study, they show that all of the three main kinds of statistical modelling can be of value for the digital art historian.

My case studies stand on their own as contributions to art history, or may complement and extend existing scholarship. In several cases they challenge accounts given by earlier authors. They show how digital methods can expose questions best addressed through conventional art historical methods. They serve as examples for other art historians interested in developing a digital approach within their own work. There are also methodological lessons for those working across the digital humanities from my work.

6.5 Looking Forward

To end this thesis I would like to suggest some of the ways my case studies might be developed within digital art history. My datasets may be of use to digital art historians. One area of work in digital art history has been the use of machine learning techniques to carry out tasks such as the attribution of the genre of a work or the gender of the artist. The datasets I built for my first and third case studies could provide input to such studies. The latter could also be of use in work which looked at the collecting policies of the art museums included in my dataset, for instance at how the careers of artists have been represented.

In my first case study I was limited in how I could use critical writing on the Paris Salon due to the small number of examples available in electronic format. As more becomes available, this would open up a number of questions which rely on a detailed analysis of critical writing as a whole for investigation. Text mining techniques such as key word searching and topic modelling could be used to look at the ways in which the changes in aesthetic value in landscape painting I read from the use of language in catalogue entries for landscape paintings were reflected in art critical discourse. They would also allow a comparison of the language used by art critics in their reviews of male and female artists.

The methodology I develop in my first case study is not limited to the Paris Salon. It could be used or adapted to examine the nineteenth-century art world in other countries and to compare developments with those in France. The British Royal Academy has published its Summer Exhibition Catalogues online, which would allow titling and other referring practices in the nineteenth-century British art world to be studied using my approach.¹⁰

In my second case study I identified some suggestive patterns connecting the price paid at auction with the age of the artist at execution and I commented this was an area meriting further investigation. Looking at the essays in auction sales catalogues and other textual resources on those artists through the use of text mining techniques would help in developing an understanding of those relationships. It might, for instance, reveal associations between the evaluative concepts used in catalogue essays and the age of the artist at execution or the career stage of the artist at the time of the auction.

I built my dataset of online catalogue metadata for modern and contemporary art over the course of 2017 and 2018. More modern and contemporary art museums will have made their collections available online since then, which could allow the analysis to be extended to include institutions in more countries. It could also encompass works created in the 2010s. Looking further forward, one question is that of whether the cyclic pattern of change in artistic interests and concerns that can be seen in my topic model continues. If the topic modelling were repeated in, say, ten years, would topic 3 remain dominant or would a topic 4 have emerged, expressing new interests and concerns or subsuming some of the interests previously expressed through topic 3?

¹⁰ The Royal Academy's Summer Exhibition Catalogues can be found at <https://www.royalacademy.org.uk/art-artists/>

APPENDIX A. STATISTICAL TESTING WITH SAMPLES

As this thesis is aimed primarily at an art historical audience my descriptions of change in artistic and critical language given in Chapter 1 avoids any references to the statistical testing of their significance.¹ With comparative measures drawn from a sample statistical testing aims to answer the question of whether those comparisons are ‘statistically significant’ in being too large to be accounted for by ordinary variation. Typically, statistical tests produce a ‘p-value’, which is an estimate of that probability. The lower the p-value, the less likely the observed difference would have arisen by chance and the greater its statistical significance. In interpreting the results of applying statistical tests researchers may split them into those that are and are not statistically significant at a particular p-value. The p-value threshold most used is 0.05, which indicates there is a less than 5% probability the results are random. Researchers may also use higher or lower levels depending upon what is appropriate for the kind of work in which they are engaged. In medical studies, for example, a p-value of 0.01 or lower may be used as researchers need to be very confident that, for instance, an expensive new drug delivers a better performance than one it is intended to replace, and the frequency of side-effects is at an acceptable level.² In marketing or product development a level as high as 0.25 may be used.³ Much less hangs on my analysis than in a medical study and I have taken a p-value of 0.1 in the case studies set out in Chapter 1 and in Chapter 2, where I also use statistical testing.

In my reading of the emergence of titling I looked at upward or downward trends over the course of time in the use of particular words or forms of catalogue entry, such as critics who did and did not use the term ‘titre’, and the use of catalogue entries to refer to more than one object. I also gave art historical interpretations to several trends in the use of words such as ‘vue’ in landscape entries, and looked at trends in the gendering of oil painting at the Salon. In these cases, an appropriate statistical test is the Cochran-Armitage test for trend.⁴ For all but one of the upwards or downwards trends presented in Charts 1.1 to 1.7 and 1.9 to 1.12 and in Tables 1.7 and 1.8 the tests confirmed what is already apparent from the chart or the data, with p-values below 0.0001. In only one case, the growth in the use of temporal or meteorological references in landscape entries over the decades from the 1840s to the 1870s,

¹ For an introduction to statistical testing with samples see Charles H. Feinstein and Mark Thomas, *Making History Count: a primer in quantitative methods for historians*, (Cambridge: Cambridge University Press, 2009), pp. 117 - 230.

² Lee Kennedy Shaffer, ‘When the Alpha is the Omega: P-Values, “Substantial Evidence”, and the 0.05 Standard at the FDA’, *Food Drug Law J*, 72(4), (2017), pp. 595 - 635.

³ Amy Gallo, ‘A Refresher on Statistical Significance’, in *HBR Guide to Data Analytics for Managers*, (Cambridge: Harvard Business Review, 2018), pp. 182 - 185.

⁴ Alan Agresti, *Categorical Data Analysis*, (Hoboken: Wiley, 2013), p. 178.

is the p-value between 0.05 and 0.1. In my reading of the emergence of titling and my comparison of male and female artists I also looked at the use of 'autre' and 'id' or 'idem' in main catalogue entries. In this case an appropriate test is the chi-square test, which confirmed the strong statistical significance of the pattern of use, with p-values of below 0.001.⁵

The other comparisons I have made in this case study are between proportions taken from the same sample, such as the proportions of entries in each generic category or of portrait entries in a particular period with different kinds of honorific, and across samples, such as the proportion of entries for portraits by male and female artists indicating a male subject. Statistical tests of such comparisons have been surveyed by the statistician George A. Seber.⁶ The statistical significance of all the comparisons I make in the main text of this chapter have been confirmed using these tests. In most cases the p-values were below 0.001, with the highest p-value of 0.06 being for a comparison of the use of explicative texts between entries by male and female artists covering the decades from the 1790s to the 1810s. Statistical testing also allowed me to identify changes that were most likely a product of my sampling process rather than genuine changes which might indicate something of art historical interest. In my analysis of portraiture I looked at the prevalence of the self-portrait. Such paintings were normally identified by the entry 'portrait de l'auteur' in the catalogue. For the portraits in my sample, there was a sharp decline in their prevalence from self-portraits representing 10% of all portraits in the 1800s to fewer than 1% in the 1820s and 1830s. However, statistical testing showed these changes were not statistically significant. A complete count of the portrait entries from the catalogues of the Salons in my dataset for the 1800s and the 1820s/1830s confirmed this result. The decline in the prevalence of the self-portrait was much more modest – from 8% in the 1800s to 6% in the 1820s and 1830s.

⁵ Agresti, (2013), pp. 75 – 77.

⁶ George A. Seber, *Statistical Models for Proportions and Probabilities*, (London: Springer, 2013), pp. 20 – 23 for comparisons in the same sample, and pp. 23 – 26 for comparisons between samples.

APPENDIX B. REGRESSION MODELLING OF AUCTION SALES DATA

B.1 Introduction

In the application of regression modelling I have used in my second case study set out in Chapter 4, the researcher looks to develop an understanding of the factors, or 'explanatory variables', determining the value of the variable of interest, or 'dependent' variable.¹ Potential explanatory variables may be identified through economic theory or the relevant literature, from empirical evidence, or from the researcher's understanding of the subject under investigation. The values of the dependent variable are related linearly to a weighted sum of the values of the explanatory variables, and the values of the weights, or the 'parameters' of the model, are determined to be those giving the best fit between the actual values of the dependent variable and the values inferred from the model.² Explanatory variables may be numerical or categorical, where the events under investigation are put into a number of classes. In my case, the dependent variable is the price achieved at auction, the numerical variables include the size of the work sold, and the categorical variables include whether the painting has a generic title or a specific title, or whether the auctioneer is one of Sotheby's or Christie's.

Faced with the same data researchers may develop different models, with different explanatory variables and with different functional relationships between those variables and the variable of interest. But for all what is key is that the model is 'valid' and enables the researcher to answer the questions posed.³ A valid model is one with a number of characteristics, each of which is tested visually and statistically before the final model is selected. Looking at the differences between the actual and predicted values of the dependent variable, these tests confirm there are no trends or other patterns in those differences which would indicate the model should be reworked. In a valid model these differences or 'residuals' are random, in the sense that residuals for different values of the dependent variable are not correlated, they follow the shape of the normal distribution or 'bell curve' and have a constant level of variation. Should any of these tests fail and the model be found to not be valid then the researcher may look to transform some or all the variables, change the explanatory factors,

¹ For an introduction to regression modelling see Simon, J. Sheather, *A Modern Approach to Regression with R*, (New York: Springer, 2010). Regression models can also be used to classify and predict rather than to explain. This approach is common in machine learning and has been used by a number of scholars working in digital and computational approaches to literary history. For one example, see Underwood, (2019).

² Sheather, (2010), p. 15 - 37.

³ Sheather, (2010), pp. 45 - 121.

exclude any data points that are anomalous and exert an undue influence upon the results and for which there is a good reason to exclude them, or use a different technique to estimate the parameters. Some of these techniques enable the researcher to correct for residuals that do not meet all the tests of randomness, and to still have a valid and usable model, but one in which there is greater uncertainty in the estimated values of the parameters than one where all the tests are satisfied. Model building is an iterative process where the researcher experiments with all these approaches to identify a valid model giving an answer to their questions.

In a valid model the researcher can be confident that the inferences drawn from the analysis are statistically sound. The results enable the researcher to determine which of the explanatory variables in the model have a statistically significant impact upon the dependent variable at any given level of significance, and the values of the associated parameters give an estimate of the scale of that impact. The model produces an estimate of the level of the statistical significance of a parameter, its p-value, and in my modelling I have used a p-value of 0.1, or equivalently a confidence level of 90%, although, as will be seen, the confidence levels associated with many of the parameters in my model are much higher. It is important to recognise that the model should be read holistically. The effect of an explanatory variable needs to be seen in the context of the other variables included in the model. The impact of a change in that variable upon the price is estimated with all the other variables held constant.

B.2 Auction Sales Data

The auction sales data I utilised in my second case study came from two sources, the art market online information and analysis portal Artprice.com and Gerhard Richter's official website.⁴ Artprice.com presents auction sales data consolidated from over 2,000 auction houses worldwide. The website can be searched by artist or keywords such as the title of a work of art, and the results filtered using criteria such as the medium, auction date, auction country and year of creation. To compile an initial long list of artists for inclusion in my modelling I queried the Artprice.com website using the title keywords 'untitled', 'abstract', 'number, and 'composition' to list artists creating works with those words in their titles in each decade from the 1950s to the 2010s.⁵ For each keyword I then identified the artists with the most auction sales in each decade (20 or more) and searched Artprice.com on their names to

⁴ Artprice.com can be found at <https://www.artprice.com/>, and Gerhard Richter's official website at <https://www.gerhard-richter.com/en/>.

⁵ Artprice automatically translates the keywords used in searches into French, German, Italian and Chinese and so my searches included works with titles in those languages.

obtain their full auction sales record. I restricted the sales record to either paintings or sculptures, to allow for like-for-like comparisons in my modelling. I then removed those artists where the total number of auction sales was less than 200 and where the works sold at auction did not include a mix of generic and specific titles, or where works with generic or specific titles were concentrated in one particular period. Artprice.com provides a thumbnail image of the object sold with all sales for which it has the image rights, and I used these to exclude artists where their works with specific and generic titles were visually distinct. As I wanted artists with an international presence in the auction market, I also removed artists where their sales were predominantly in one country, or where the average sales price was below \$50,000.

For one artist, Gerhard Richter, I used the sales data given on his official website for those of his paintings classified as abstracts as it is a more consistent data source than Artprice.com. Although the paintings listed as abstracts go back to 1964, I restricted my analysis to those created from 1976 as the first year in which Richter began to use the title 'Abstraktes Bild' and also the year when several scholars have identified a shift in his practice to engage with abstraction.⁶ I also excluded the remainder of his paintings which are classified under various subject headings on his website such as landscapes, skulls, and buildings as each group is visually distinct. For each painting the website presents an image, if available, the title in German and in English translation, the official catalogue number and various characteristics of the painting itself. If a work has been presented at auction, the website gives the full sales data, with estimates and sales figures presented in the local currency and US Dollars. The large majority of sales figures include the buyer's premium, the commission paid by the buyer to the auction house, some record the hammer price, and a small number do not say what the sales figure represents.⁷ For my modelling I compiled the data for all of the auction sales where the buyer's premium is included in the reported sales price. For the painting I included its title, dimensions, year of execution, medium and support. For the sale I included the price in US Dollars, the auction house, location and date. Using the sales data in US Dollars allows for direct comparisons to be made between sales at different locations.

For the other artists I used the sales data given by Artprice.com. This is largely the same as that given on Richter's website except in some cases an estimated year of creation or a range of years is given, and all prices are the hammer price. As the buyer's premium is significant,

⁶ See, for instance, the curator Robert Storr's catalogue essay in Richard Storr, (ed.), *Gerhard Richter: Forty Years of Painting*, (New York: The Museum of Modern Art, 2002), pp. 68 - 69.

⁷ The full cost to the buyer will also include any local sales taxes and other charges such as resale royalties with sales in the European Union.

typically 10% or more of the hammer price, this should be borne in mind when comparing sales figures for Richter and other artists.

I will come on next to describe my model development process, but for some artists I was unable to develop a valid model, which may reflect there being significant drivers of the prices realised at auction which are not associated with any of the characteristics I have included in my model. This narrowed my long list down considerably and left me with the twelve artists I have looked at in this case study, for eleven of which I used the sales data on paintings and for one on sculptures.

For the twelve artists I have modelled the data comes from sales at 133 auction houses in 24 countries. It is dominated by sales at Christie's or Sotheby's, which handled over 70% of all sales, and auctions in the United States or the United Kingdom, which were the location for around 70% of all sales.

B.3 Regression Model Development

To prepare the sales data for my modelling I had to categorise the non-numerical data and decide how to measure the numerical data. I also had to do some cleaning and processing of the Artprice.com data. Artprice.com consolidates its data from a large number of auction houses and it soon became apparent to me as I worked on their data that the same painting may not be described the same way by different auction houses nor by the same auction house for sales at different times. There are also some significant gaps in the data. As part of the iterative process of model building I also had to return to the Artprice.com data and make some further changes.

In building a regression model the researcher has to specify the functional relationships between the prices paid at auction and the explanatory variables included in my model. The sales prices for most artists had a very skewed distribution, with a small proportion of works commanding prices very much higher than the average. In such circumstances it is standard to adjust for this skewness and use the natural logarithm of the variable in the model rather than the actual value, and I found this approach worked well for my data. I will now discuss each of the categorical and numerical explanatory variables and set out how they have been represented in my model.

For the categorical variables in the model the standard approach that I followed is to have what is called a 'dummy' variable taking on the value 1 if the sale is in the category and 0 if it

is not.⁸ So, for instance, if the sale is held in the United States the corresponding dummy variable has value 1, if it is held elsewhere the value is 0. The titles of works were categorised as generic or specific. In many cases the Artprice.com sales data does not give the support of the work, and so I did not include that data in my modelling. For the medium, there are some inconsistencies in the Artprice.com data with, for instance, the same work reported as being executed in oil and as being in mixed media. My initial modelling indicated that paintings where the medium was recorded in the Artprice.com data as oil or as oil and other media commanded a premium for several artists compared with other paintings. For that reason, I decided to retain medium in my modelling, counting oil or oil and other media as one category and all other media as another. Almost all the sales of paintings by Gerhard Richter and Joan Mitchell in my dataset are of works executed in oil whereas almost all the sales of paintings by Christopher Wool are not, and so execution in oil was not a relevant variable for those artists. However, the results need to be interpreted with a degree of caution since some of the paintings categorised as in other media are likely to have been executed in oil, and so the sales premium or discount attached to paintings in oil compared with other media may have been underestimated.

With the auction house, I grouped together Sotheby's and Christie's as one category and all other auction houses as another. With my model for each artist, I only included the locations with a significant number of sales as separate categories, grouping the remainder together as another. The resulting locations included in the models were the United States, the United Kingdom, Hong Kong and the native country of the artist. Almost all of Christopher Wool's auction sales are in the United States or the United Kingdom and so I could only retain one in my model and chose to keep the United States. Of the artists I have modelled, only Yashimoto Nara had sufficient sales in Hong Kong for that location to appear in their model, and Joan Mitchell had sufficient sales in France. As the United States was included as a sale location for all artists with significant sales in that country, I did not include a native country variable for American artists. For Albert Oehlen there were insufficient sales in Germany to include the native country as a variable in my model for his sales.

To test the robustness of this approach to classifying the auction house and location of the sale I also looked at an alternative version of my model been used in other studies in which they are combined. I included sales at Sotheby's or Christie's in the United States as one category, and sales at Sotheby's or Christie's in the United Kingdom as another. Other sales

⁸ Sheather, (2010), pp. 31 - 33.

in the United States or the United Kingdom was included as one category, as were sales in the artist's native country and in other locations with sufficient sales volumes. The results from this version of the model were very similar to those I have presented in Chapter 4.

For the numerical data, I wanted to be able to look at average sales prices at different times as well as in different locations and so I converted the actual US Dollar sales prices into real values, that is adjusting for inflation using the Consumer Price Index. The baseline month used was June 2019, the most recent month for which I have collected data.

The Gerhard Richter website provides a single year of creation for each of his abstracts. Where a definite year of creation was not given in the Artprice.com data I took the estimated year or the mid-point of a range of years. For my modelling I converted the year of creation into the age of the artist at that time. As I wanted to look at how collectors value paintings produced at different times of an artist's career, I added the square of the age to my model, the most common approach to this sort of question used in the literature. This allows for models where the price of works stays roughly constant, falls with the artist's age, rises with their age, or has a period where the price is at a peak or a trough.

To look at how each artist's prices have changed over time I included the month of the auction in my models, with the variable taking on the values 1 for January 1984, 2 for February 1984 and so on. This allows an estimate to be made of the underlying or long-term rate of appreciation in the auction value of an artist's paintings. It is more common in the literature to use the year of the auction for this purpose, but in practice the two approaches give very similar results.⁹ The late 1980s was a boom time for the contemporary art market and in my initial modelling I found that for artists with sales during that period my models were consistently under-predicting auction prices. I corrected for these boom years adding a dummy variable which took on the value 1 for sales in 1988 and 1989 and 0 for sales in other years in my model. The contemporary art market also experienced booms years in 2008 and 2009 followed by a bust, however I found that corrections for these years were not statistically significant and so I have not included them in my model. I also found that for two artists, Sam Francis and Asger Jorn, their sales prices were over-predicted in the early and mid-years of the 1990s and I added additional dummy variables to my model for these lows. For those two artists from 1997 and, for all the other artists bar one, from 1991 the model did not exhibit any

⁹ This variable is also often interpreted in terms of a rate of return to investors in an artist, however that is misleading as it ignores all the costs associated with buying and owning an artwork including the buyer's premiums, sales and other taxes, and storage and insurance costs, all of which can be substantial.

consistent over- or under-predictions, indicating that their prices appreciated consistently at around the long-term rate over that period. The exception was A. R. Penck, whose prices at auction have been more volatile than those of other artists and in addition to a boom in the late 1980s, and a slump from the mid-1990s to the mid-2000s, prices have boomed again in recent years. I corrected for this volatility in my model for A. R. Penck.

To look in more detail at the underlying trends in price I looked at an alternative and more complicated form of the model where a price index is created for each artist. Rather than having a long-term rate of appreciation coupled with corrections for boom and busts, in this type of model sales in each year for each artist are associated with a dummy variable. So, for instance there is one dummy variable taking on the value 1 for sales in 2001 and 0 for sales in other years, and another taking on the value 1 for sales in 2002 and 0 for sales in other years. Charting the parameter values associated with these variables gives a measure of how underlying prices have been changing for each artist. This confirmed the picture given by the simpler model which, for ease of presentation and understanding, I have retained as the model I have based my analysis upon.

With paintings I combined the height and width given in the sales data into one measure of area. The differences in reported size for sales of the same painting in the Artprice.com data could be substantial, and the performance of my model was improved once I had made some corrections for these differences. Where I could be confident through examination of the image, title and year of creation, that my dataset included multiple sales of the same painting I took the most commonly given or, if not available, the average of the sizes for use in my modelling. For most of the artists I modelled the size of their paintings had a skewed distribution with a small number of very large works, and taking the natural logarithm of the size resulted in a better model than using the size itself.

As I have discussed in Chapter 4, Alexander Calder made wire sculptures and works that he called 'mobiles' and 'stabiles'. My initial modelling indicated that sculptures recorded as mobiles commanded a premium, all other factors being equal, compared with other sculptures. However, examination of the Artprice.com data where I could be confident the same work was being sold showed they were not always reported as mobiles. For this reason, I decided to restrict my modelling to sculptures identified as mobiles in the Artprice.com data. As articulated, moving objects there are no fixed dimensions to a Calder mobile, and the sizes given in the Artprice.com data vary significantly. For some sales three dimensions are given and in others two. For my modelling I decided to take the longest dimension reported in the

data for each sale and made further adjustments for the size on the same basis as my adjustments to the size for paintings, using the natural logarithm in my model. Despite these uncertainties my model for sales of Calder's mobiles delivers a similar level of explanatory performance to my models for other artists.

In several studies, researchers have looked at whether there is a maximum size that appeals to collectors and beyond which the price paid at auction declines on average, all other factors being equal. I also investigated this question by adding the size to my model alongside the natural logarithm of the size, a form for the model which, as with including the artist's age and its square, allows for a size at which prices peak or trough, if there is one, to be identified. I found that for only two artists was there a maximum size included in the range of sizes of their paintings sold at auction. For the other artists either there was no maximum size, or the maximum size was much greater than that of any of their works sold at auction. For these reasons I have retained the simpler model with the natural logarithm of size only included.

I have already discussed some of the adjustments I had to make to the sales data as part of the iterative process of model building. In addition, I excluded any sales in the Artprice.com data for which information was not provided on the size or the year of creation, as both are significant explanatory variables in my model. To end up with valid models for each artist I also had to make some further changes to the data which involved removing sales that had a significant impact upon the tests for validity and were 'outliers' in being substantially under- or over-predicted by the model. For A. R. Penck I found that sales of his paintings at a Berlin auction house were consistently over-predicted by my model and so I removed those sales. For Sigmar Polke the model consistently over-predicted sales of a group of paintings with the same title. Investigation of these sales online identified that these were editions, and so I removed them from my dataset. With Alexander Calder the model performance was significantly affected by the sales of a group of his mobiles with the same or similar title, whose sales prices were over-predicted by my model. Researching these works online identified they are among the mobiles that have been most prone to forgery. The over-prediction by my model indicates collectors are aware of these risks and discount the price they are prepared to pay for these works compared with other mobiles. I excluded sales of these mobiles from my model. For Asger Jorn, my model consistently over-predicted paintings executed in the late 1930s, a period when he spent much of his time studying and working in Paris. I excluded sales of these works from my model. Finally, for several artists I excluded a few outlier sales which were substantially under- or over-predicted by my model and affected the model's validity. These may represent sales where the demand for the artist's work was low or where

a small number of collectors were bidding aggressively for a work. For instance, *Summer #1* painted by Sam Francis in 1953 sold at auction in the United States in 2016 for \$10.4 million, nearly twice the next-highest price achieved at auction by Francis at the sales included in my dataset. These substantial over- and under-predictions may also represent errors in the sales data such as misreporting of the dimensions of the painting or not reporting a painting as an edition.

The model I developed for each artist has the general form standardly used in the cultural economics literature I have surveyed in Chapter 4. It was as follows:

$$\ln(P_i) = c + \beta_1 \ln(S_i) + \beta_2 O_i + \beta_3 SC_i + \beta_4 US_i + \beta_5 UK_i + \beta_6 HK_i + \beta_7 F_i + \beta_8 N_i + \beta_9 M_i + \beta_{10} T_i + \beta_{11} A_i + \beta_{12} A_i^2 + \beta_{13} B_{1i} + \beta_{14} B_{2i} + \beta_{15} B_{3i}$$

Where \ln is the natural logarithm and the subscript i labels each of the sales for the particular artist. P is the price in real US dollars of the work of art sold. The constant c and β_1 to β_{15} are the parameters of the model whose values are determined by the linear regression modelling. S is the size of the work sold. O is the dummy variable for works executed in oil. SC is the dummy variable representing sales at Sotheby's or Christie's. US , UK , HK , F and N are the dummy variables for sales in the USA, the United Kingdom, Hong Kong, France and the artist's native country respectively. M is the month of the sale. T is the dummy variable for sales of works with a specific title. A is the age of the artist at execution. B_1 is the dummy for the art market boom in the late 1980s. B_2 is a dummy for the slump in prices in the mid 1990s in the models for Sam Francis, Asger Jorn and A. R. Penck. B_3 is the dummy for the boom in prices in the late 2010s in the model for A. R. Penck.

For all artists I applied the standard statistical and visual tests for model validity. Another potential issue with regression models is 'collinearity', where two or more of the variables are closely aligned. In such cases the estimates of the parameters associated with those variables can be very sensitive to small changes in the data and interpretation of those parameters is problematic. The residual tests used were the Durbin Watson test for autocorrelation, the Jarque-Bera and Shapiro tests for normality, and the Breusch-Pagan test for heteroskedasticity. I also used the variance inflation factor tests for multi-collinearity. In the visual tests, all the explanatory variables were charted against the standardised residuals. For five of the artists the model passed all the tests, however for the others it passed all bar the Breusch-Pagan test for constant variation in the residuals. Faced with such a model, researchers often use a 'heteroskedastic consistent' modified estimation technique which is robust to that non-constancy, one consequence of which is that there is less confidence in the

estimates of the parameters than in standard linear regression. I have used the version of this modified technique, 'HC3', which is typically recommended in the literature, and, to allow for comparison across all artists of the estimated values of the parameters, have applied it to all twelve artists.¹⁰

B.4 Regression Modelling Results

Table B.1 presents the results of my regression modelling. For each artist it gives the values of the parameters associated with each of the explanatory variables. The figure in brackets below the parameter value is the 't-value', a measure of the uncertainty in the estimated value of the parameter. The asterisks after each parameter value indicate the statistical significance of the parameter estimate, or, correspondingly, the level of confidence that the explanatory factor is an influence on the sales price. Three stars indicate p-values of 1% or less, or equivalently significance at a confidence level of 99% or higher, two stars 95% or higher, and one star 90% or higher, which is the confidence level I have adopted in my interpretation of the model results. No stars indicates that the parameter is not significant at the 90% level. In the last row I present what is called the 'R squared' statistic, which is the proportion of the variation in the natural logarithm of the price achieved at auction for each artist that can be explained by my model.

¹⁰ J. Scott Long and Laurie H. Ervin, 'Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model', *The American Statistician*, 54/3, (2000), pp. 217 - 224.

	Alexander Calder	Sam Francis	Asger Jorn	Martin Kippenberger
Constant	8.0535*** (40.632)	10.1204*** (21.015)	4.2193*** (7.782)	11.446*** (6.035)
Log of Size	0.8012*** (41.435)	0.5559*** (24.296)	0.7736*** (0.7736)	0.9613*** (18.968)
Oil	n/a	0.5799*** (5.152)	0.8844*** (13.503)	0.7746*** (6.996)
Sale at Sotheby's or Christie's	0.1351* (1.854)	0.3086*** (3.268)	0.0536 (0.800)	0.2375* (1.648)
Sale in the United States	0.1645* (1.897)	-0.1568 (-1.633)	n/a	0.5547** (0.4762)
Sale in the United Kingdom	0.1672* (1.792)	-0.2587** (-2.304)	0.0811 (1.389)	0.3329* (1.675)
Sale in Hong Kong	n/a	n/a	n/a	n/a
Sale in France	n/a	n/a	n/a	n/a
Sale in native country	n/a	n/a	0.0645 (1.016)	0.2437 (1.130)
Auction date, month	0.0071*** (45.264)	0.0020*** (6.538)	0.0008*** (4.283)	0.0064*** (7.744)
Specific title	0.1051** (41.431)	0.3877*** (4.815)	0.1993*** (4.023)	0.1945* (1.651)
Artist's age at execution of work	-0.0240 (-0.989)	-0.1708*** (-5.385)	-0.0387* (-2.555)	-0.6894*** (0.117)
Artist's age squared	-0.000025 (-0.151)	0.0015*** (5.257)	0.00050* (1.749)	0.0097*** (5.642)
R squared	0.821	0.750	0.701	0.754

Table B.1. Results from Regression Modelling

	Joan Mitchell	Yoshimoto Nara	Albert Oehlen	A. R. Penck
Constant	2.6935*** (3.888)	-10.3326*** (-7.352)	-4.0182*** (-3.518)	7.9336*** (11.198)
Log of Size	0.7455*** (24.433)	0.7157*** (26.698)	0.6173*** (13.332)	0.5541*** (18.737)
Oil	n/a	0.2390*** (2.632)	0.0396 (0.923)	0.2060*** (2.970)
Sale at Sotheby's or Christie's	0.1390 (1.172)	0.0586 (0.642)	0.1937* (1.952)	0.1427 (1.455)
Sale in the United States	-0.1859 (-1.431)	-0.0527 (-0.412)	0.3629** (2.426)	-0.0367 (-0.282)
Sale in the United Kingdom	n/a	-0.0237 (-0.165)	0.3754 (2.464)	0.1453 (1.476)
Sale in Hong Kong	n/a	0.1261 (0.961)	n/a	n/a
Sale in France	-0.2501 (-1.583)	n/a	n/a	n/a
Sale in native country	n/a	-0.4043*** (-3.398)	n/a	-0.1631** (-2.103)
Auction date, month	0.0090*** (27.166)	0.0094*** (11.843)	0.0118*** (15.877)	0.0002 (0.648)
Specific title	-0.10197 (-0.250)	0.4722*** (0.4722)	0.2289** (2.176)	0.1360** (2.158)
Artist's age at execution of work	0.0213 (0.833)	0.5572*** (8.414)	0.2282*** (4.072)	-0.1205*** (-4.702)
Artist's age squared	-0.0004 (-1.339)	-0.0064*** (-7.599)	-0.0028*** (-4.306)	0.0012*** (4.633)
R squared	0.863	0.794	0.805	0.601

Table B.1. Continued.

	Sigmar Polke	Gerhard Richter	Cy Twombly	Christopher Wool
Constant	5.1336*** (6.751)	-5.9257*** (-3.356)	3.1213** (2.328)	-14.4434*** (-5.617)
Log of Size	0.8743*** (21.335)	0.8676*** (38.352)	0.8428*** (14.393)	0.7762*** (12.189)
Oil	-0.0567 (-0.385)	n/a	-0.1069 (-0.783)	n/a
Sale at Sotheby's or Christie's	0.1300 (0.817)	0.2126** (0.103)	0.1300 (0.389)	0.2429* (1.930)
Sale in the United States	0.7189*** (3.600)	0.5693*** (0.181)	0.8674** (1.974)	0.0692 (0.547)
Sale in the United Kingdom	0.8207*** (4.059)	0.5261*** (2.920)	0.7669* (1.768)	n/a
Sale in Hong Kong	n/a	n/a	n/a	n/a
Sale in native country	0.3261* (1.651)	0.4869** (2.314)	n/a	n/a
Auction date, month	0.0075*** (15.470)	0.0127*** (41.529)	0.0085*** (15.276)	0.0155*** (21.163)
Specific title	0.2432*** (2.714)	0.0382 (0.551)	-0.4424*** (-3.439)	0.4025*** (2.813)
Artist's age at execution of work	-0.1727*** (-6.005)	0.3348*** (5.406)	-0.0255 (-0.572)	0.6568*** (5.109)
Artist's age squared	0.0014*** (4.64)	-0.0027*** (-5.024)	0.0001 (0.729)	-0.0080*** (-5.102)
R squared	0.704	0.889	0.712	0.714

Table B.1. Continued.

To interpret the values of the parameters in regression models researchers typically take a confidence level and look to read the parameters that are statistically significant at that level or better, and may not give a reading to the other parameters. I have taken a p-value of 0.1 in my modelling, as can be seen, the level of statistical significance associated with many of the parameters in my model is much higher. To interpret each of the parameters you turn to the model formula and consider how a change in the value of the associated explanatory variable feeds through into a change in the average price achieved at auction, all other explanatory variables held constant. I will now do that for an example of each of the explanatory factors to show how the figures given in Table 2.2 relate to those in Table B.1.

In my models, the size of the painting or mobile is a significant factor for all artists. As the model uses the natural logarithm of size the interpretation of the parameter is not the same as it is with the other variables. The form of the model means that, whatever the size, a given proportional increase in that size results in the same proportional increase in the price. With Ager Jorn, for example, a doubling of the size of a painting, on average and with all other

factors held constant, was associated with a proportional increase in the price of $2^{(0.7736)} - 1$, or 71.0% in percentage terms, as given in Table 2.2.

The categorical explanatory variables in my models take on the values 0 if the sale or work is not in the category and 1 if it is. The price change associated with the variable taking on the value 1 compared with 0 is given by exponentiation, the inverse of the natural logarithm. A painting executed by Yoshimoto Nara in oil sold, on average and all other factors being equal, for $e^{(0.239)} - 1$, or 27.0% more than a painting in mixed media or acrylic, the other media included in my data for the artist. For Martin Kippenberger, on average and with all other factors held constant, a painting sold at Sotheby's or Christie's for $e^{(0.2375)} - 1$, or 26.8%, more than a painting sold at another auction house. A painting by Sigmar Polke sold, on average and all other factors held constant, for $e^{(0.7189)} - 1$, or 105.2% more, at an auction in the United States than one sold at an auction in one of the countries not explicitly included in my models, which in this case are Austria, France, China, Italy and The Netherlands.

For all the artists in my models the date of auction is a significant factor. With Alexander Calder, for example, a mobile sold at auction in one month would, on average and with all other factors held constant, sell for $e^{(0.0071)}$ times the price it would have achieved in the previous month. To get an annual rate of appreciation you compound this monthly growth rate, and in percentage terms this is $e^{(12 \times 0.0071)} - 1$, or 8.9% as given in Table 2.2. For Albert Oehlen, on average, paintings with specific titles sold for $e^{(0.2289)}$, times the price of mobiles with generic titles. As a percentage premium this is $e^{(0.2289)} - 1$ or 25.7% as presented in Table 2.2.

The model formula for how the average price of a Sam Francis painting changes with his age when it was executed is $\text{price} = e^{(-0.1708 \times \text{age} + 0.0015 \times \text{age} \times \text{age})}$. This is charted in Chart B.1. The curve should be interpreted as a stylised profile, giving the underlying trend in the price of Francis' paintings with his age at creation. In addition, the absolute level of the price has no meaning, rather, what has meaning is the relative levels at different ages and for ease of understanding I have scaled the level to have a value of 1 for the earliest age at creation appearing in the auction sales data.

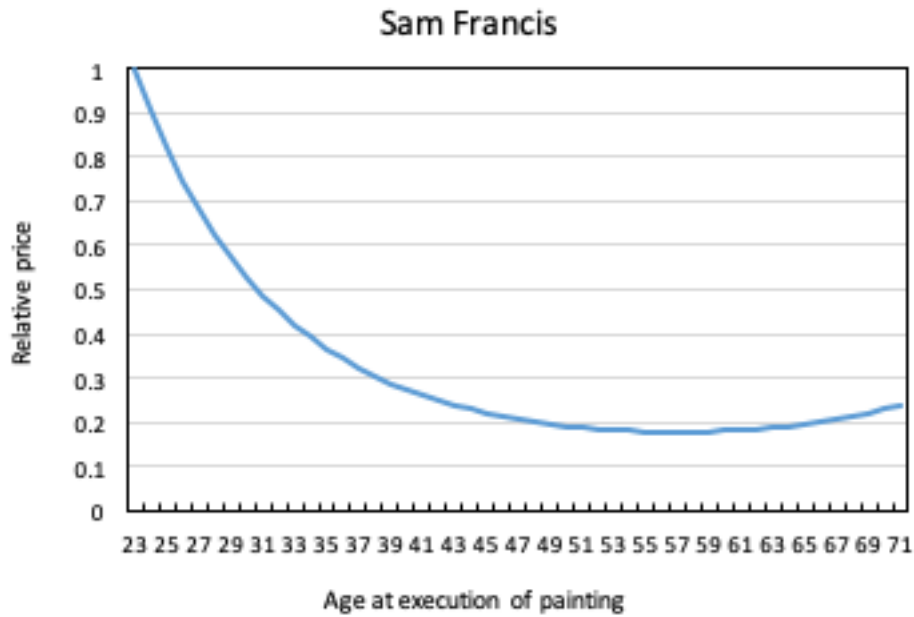


Chart B.1. Relative prices for paintings executed at different ages by Sam Francis.

Finally, the R squared statistic measures the correlation between the predicted and actual prices, or the proportion of the variation in the natural logarithm of the sales price that can be explained by my model. In regression modelling, adding a variable to a model automatically increases the R squared statistic and researchers often use an alternative R squared measure that adjusts for that property. However, in my case the number of variables is small compared to the number of auction sales I have modelled, and the two statistics are very close in value. For ease of understanding, I have retained the R squared measure.

APPENDIX C. CORRESPONDENCE ANALYSIS, TOPIC MODELLING AND MODEL DEVELOPMENT, AND PARTS-OF-SPEECH TAGGING

C.1 Correspondence Analysis

In Chapter 3 the strong relationship in my sample of Salon criticism between the period in which the criticism was written and the use or not of the term 'titre' was a key component of my analysis of the emergence and adoption of titling. As the use or not of the term 'titre' is a binary category it was straightforward for me to use a graphical presentation to identify that relationship. With the dataset of metadata I have used in Chapter 5 I similarly wanted to see if there were any significant relationships between the words used in titles and the decades in which the works of art they name were created. In this case the number of elements in these two ways of categorising the data is very large with twelve decades and around 28,000 words appearing in titles at one time or another and so a simple graphical analysis was not appropriate. It would not have been feasible, for instance, for me to examine 28,000 charts showing the levels of use of each word in each decade to identify any clusters of words which tended to be used together. And so I turned to the sophisticated descriptive statistical technique of correspondence analysis, which was developed to deal with this kind of question.¹

Correspondence analysis starts from a table whose rows and columns represent different ways of categorising a dataset. For instance, in my case, each row would be a decade and each column one of the 28,000 words used in titles. Each entry in the table gives a frequency count, in my case how often a particular word was used in the titles of works created in a particular decade. So, each row in the table is a profile of the level of use of each of the words in a particular decade, and each column a profile of the levels of use of a particular word across all decades. Correspondence analysis gives a way of graphically presenting such row and column profiles that allows relationships between them to be read off. The graph is constructed in such a way that there is an approximate relationship between the geometrical distances separating profiles and the extent to which they are similar to each other.² It is not

¹ For a detailed exposition of correspondence analysis see Michael Greenacre, *Correspondence Analysis in Practice*, (2nd ed.), (London and New York: Chapman and Hall, 2007).

² The measure of similarity used in correspondence analysis is 'chi-square distance'. To calculate this measure for row profiles, each row is transformed into percentages and each element of the transformed row is weighted down by the overall frequency of that element in the table. These scalings ensure that the distance measures are not dominated by the most frequent components. Chi-square distance is the normal geometric distance between these re-scaled profiles. The chi-square distance between column profiles is calculated analogously. The chi-square statistic for the data tabulation can be related to these chi-square distances.

possible to present all the variation between profiles on the two dimensions of a graph, but the mathematics of correspondence analysis determines how much of the total variation it captures, and so, how good the approximation is.

There are three ways in which the results of correspondence analysis are presented and interpreted.³ In the first, the profiles for one category are plotted. The level of similarity between these profiles can be read from the chart, as similar profiles are close together, and it is common to see if they can be clustered into a small number of groups. For example, in my case, the decades can be plotted and compared or clustered. The centre or 'origin' of the chart also has interpretive significance as it represents the 'average' profile. In my case, a fictitious decade in which each of the 28,000 words (or the 'vocabulary') is used equally frequently. The further decade profiles are from the origin the more discriminating they are as the use of words in the titles of those decades differs more from the average. To aid in interpretation it is usual to also add to this plot points corresponding to fictitious 'extreme' profiles, each of which is concentrated in one element of the other category. So, in my case, one point for each word in the vocabulary representing the extreme case of a decade where all the titles consist of that word alone (or, to avoid too many points, the extreme profiles for the most frequently used words). For each word, the decades on the same side of the origin as the point corresponding to its extreme profile are those in which that word is used more frequently than average. The further along the line from the origin to the extreme profile point a decade is, the more heavily that word is being used in that decade compared with others. Correspondingly decades on the opposite side of the origin are those where the word is used less frequently, and those furthest away are the decades in which it is used the least.

In the second graphical presentation, the corresponding analysis is presented for the other category. In my case, this presentation allows words to be compared and clustered, and the inter-relationships between word and decade to be explored through plotting extreme points for each decade.

It is a consequence of the mathematics of correspondence analysis that the total variation captured in the graphical presentations of the two categories are equal. The two separate plots can therefore be combined into a single plot, and in this case the points corresponding to the fictitious extreme profiles are not displayed. This 'symmetrical' plot is the one most

³ I have adapted these principles from those given in Greenacre, (2007), and by the data analysts Phillip M. Yelland, <http://www.mathematica-journal.com/2010/09/an-introduction-to-correspondence-analysis/>, and Tim Bock, <https://www.displayr.com/interpret-correspondence-analysis-plots-probably-isnt-way-think/>.

commonly used, and is the plot presented by the Voyant correspondence analysis tool. In this plot the similarities, differences, and clustering of the elements for each category separately can be read off. It may, however, be misleading to examine the relationships between categories through the points lying on the same or opposite sides of the origin as they may not be close to the extreme points, which are not plotted. However, in practice, symmetrical plots are often interpreted in this way. As the statistician Michael Greenacre discusses, the distortions introduced through treating the points representing the actual data as extreme points are often not great, although it is always advisable to return to the original data or to other ways of presenting it to confirm any inter-relationships suggested by the correspondence analysis.⁴

C.2 Topic Modelling and Model Development

Topic modelling is a predictive statistical text-mining technique that attempts to capture the intuition that collections of documents can be organised and annotated by topics or themes running through them, as expressed in the words used in them.⁵ Technically, it does this through modelling the documents as having been generated from a number of 'topics' by an underlying probabilistic process. In the topic model, each document is represented as a probability distribution over all the topics, so document 1 will be modelled as x% topic 1, y% topic 2 and so on for however many topics there are in the model, with the percentages summing to 100%. Each topic is a probability distribution over the 'vocabulary' - all the words used in all the documents - so topic 1 is x% word 1, y% word 2 and so on for all the words in the vocabulary. To generate a model of a document, you begin by picking a distribution over topics and then generate a word by first randomly choosing a topic from that distribution, and then randomly selecting a word from that topic according to its distribution over the vocabulary. The process of word generation is repeated as many times as there are words in the document to be modelled. It is then repeated for each of the original documents. The topic modelling algorithm works by varying the distributions to maximise the probability the generated documents will contain the same words as the original documents being modelled. In comparing modelled and original documents no account is taken of the order of words in them - each is treated as a 'bag of words' where what matters is only their frequency of occurrence.

⁴ Greenacre, (2007), p. 267.

⁵ For introductions to topic modelling see David M. Blei, Andrew Y. Ng, and Michael I. Jordan, 'Latent Dirichlet Allocation', *Journal of Machine Learning Research*, 3, (2003), pp. 993 – 1022, David M. Blei, 'Probabilistic Topic Models', *Communications of the ACM*, 55/4, (2012), pp. 77 – 85, and Jordan Boyd-Garber, Yuening Hu and David Mimno, 'Applications of Topic Models', *Foundations and Trends in Information Retrieval*, XX/XX, (2017), pp. 1 - 154.

Topic modeling is a form of unsupervised machine learning in the sense that the researcher selects the documents and the values of the parameters within which the analysis is to proceed, but it is the algorithm that then determines the topics that best generate those documents. There is no independent check of the validity of those topics. However, and as my use of the technique in Chapter 5 makes clear, through the choices made by the researcher and through their interpretation of the results there is a substantial amount of 'subjective' input into topic modelling.

The implementation of topic modelling I utilised is that developed by Andrew McCallum as part of his MALLET toolkit, and in what follows I describe topic modelling using his tool.⁶ Other implementations of topic modelling take a similar approach. With MALLET there are two main steps in the development of a topic model. In the first step the researcher specifies the documents to be modelled and provides a 'stop list' of words that are not of analytical interest and are excluded from the analysis. Standardly, these are function words whose role is primarily syntactic, and they include the conjunctions (such as 'as', 'and', 'but', 'for' and 'so') and the definite and indefinite articles. Function words are also typically very common and so would dominate the topics in the model if retained, making those topics harder to interpret. The researcher additionally defines how the software is to 'tokenise' or divide the documents into word tokens, for instance whether to retain punctuation marks. Keeping punctuation marks would result in the word 'token' and the character string 'token!' being regarded by the software as tokens of distinct word types. Such a distinction might be of interest to a researcher investigating an author's style of writing including their use of exclamation marks. MALLET takes the documents selected by the researcher, extracts all the given list of stop words and tokenises the texts appropriately, treating each as an unordered collection of tokens. The 'bags of words' output from this step then forms the input to the development of the topic model itself.

Researchers often exclude proper names before tokenising. There are a number of applications that aim to recognise named entities in a text and so can be used to enable their extraction. In my case I investigated the use of the state-of-the-art named entity recogniser developed by the Stanford University Natural Language Processing Group.⁷ However, I found

⁶ The MALLET toolkit can be found at <https://mimno.github.io/Mallet/index>. As a computer-based processing technique topic modelling is not restricted to the analysis of documents. It can be used to analyse any collection of documents or data that can be digitised and tokenised, such as images or musical scores.

⁷ The Stanford University Natural Language Processing Group named entity recogniser (NER) software can be found at <https://nlp.stanford.edu/software/CRF-NER.html>.

it did not work well with titles. Titles are short texts providing very little linguistic context to the application, and many words in titles have their initial letters in upper case, which is one of the features the application uses to identify likely names. The application missed some names and mis-classified as names a significant number of words that were not. As a result, I decided to retain proper names in my titles.

Another common transformation of the documents to be topic modelled is ‘stemming’ in which words are reduced to their root form, for instance the ‘s’s are removed from plurals and ‘walk’, ‘walked’ and ‘walking’ are all rendered as ‘walk’. I developed my topic model using both stemmed and unstemmed documents, and then compared the results. The two models were very similar and for ease of understanding all the models I describe in Chapter 5 are derived from the unstemmed documents.

In my modelling I considered four ways of combining titles together as the input documents for the topic modelling. In each case I excluded punctuation marks other than the hyphen, which I retained so that titles such as ‘self-portrait’ were retained as one token, and other non-alphanumeric symbols when tokenising these texts and used the standard list of English stop words provided with MALLET. I will look first at my modelling of titles combined by the decade in which they were used, which is the model described in Chapter 5, before coming on to look at the three other ways.

The second step in the use of MALLET is to run the algorithm that models the ‘bags of words’ output from the first step. There are a number of different parameters whose values can be varied by the researcher. In my case I looked first at whether to have ‘hyper-parameter’ optimisation, which determines whether all topics are to be given equal prominence across the collection of texts as a whole, or whether the relative importance of topics in the model can be different. As I experimented with this option for models with different numbers of topics, the first point I noted was that having hyper-parameter optimisation resulted in models that were typically more discriminative than if topics were given equal prominence, in the sense that the distributions of topics in the model for the titles used in each decade were more concentrated in a smaller number of topics. Such models are more likely to be interpretable than those where the topic distributions are more spread out across the range of topics. The results presented in Chapter 5 all come from topic modelling with hyper-parameter optimisation.

I then looked in detail at the consequences of varying the number of topics in the model. At one extreme I set the number of topics to be three, forcing the algorithm to look for any high-level trends in the data. At the other extreme I set the number of topics to be 100, which allowed the software to develop a fine-grained model and look for topics of importance only in specific periods. I also looked between these two extremes at models with 5, 10, 20, 30 and 50 topics. MALLET is a non-deterministic tool for topic modelling in that running the algorithm twice with the same input texts and parameter values produces similar, but not necessarily identical results. For each number of topics I ran the algorithm multiple times to ensure the results I would then interpret were not outliers. The model presented and interpreted in Chapter 5 should be seen in that light. Whatever the number of topics, the same or a very similar pattern emerged with most of the models. Typically, in each of the models there were three long-term prominent topics, with each topic the most prominent in a particular period within an overlapping pattern of change through the decades from the 1900s to the 2010s. The most important words in each of the three topics were closely related to the three clusters revealed through my correspondence analysis.

As expected, as I increased the number of topics the models became better at isolating those words whose use was concentrated during a short period of time. Their importance diminished in the long-term topics, and they typically became among the highest-weighted words in others. For example, the subjects often favoured by the Cubists such as guitars, bottles and violins were given high weights in topics of prominence in the 1910s and 1920s but no other decades. However, with more than 20 topics the model included many redundant topics with a weight below 5% across all decades and so contributing little to generating the words used in the titles. In fact, only around 20 topics in each model were of significance. For instance, the average number of redundant topics in the 30 topic models was 9, and in the 50 topic models was 28. Such topics would be largely irrelevant to any art historical interpretation of the model. With 50 and 100 topics, the three long-term topics themselves were beginning to fragment. In particular, these models often had two topics equally prominent in the early twentieth century and sharing several important words. In consequence their distinctiveness in the model was reduced, and correspondingly their interpretability. From the range of topic numbers I examined 20 provided the optimum balance between having distinctive long-term topics and all or almost all of the other topics being of short-term relevance and discriminating between decades.

It is common in topic modelling to work with units of text much smaller than the documents I used, for instance scholars may split a document into its paragraphs or into parcels of 500 words. Researchers find that modelling these smaller units delivers models that are more interpretable.⁸ In my case I looked at three other ways of combining titles into smaller textual units. Firstly, I output each title as a separate document, and so used MALLET with 63,531 separate documents. With a small number of topics, the results were very unstable and heterogeneous each time I ran the model, but as I increased the number of topics the model began to identify groups of words that often appeared together in titles. For instance, the most important words in one topic included the words 'portrait', 'sitter', 'Mr' and 'Mrs', another the common subjects of still-life paintings, and a third the word 'number' and the numerals 1 to 6. Whilst giving an understanding of how the levels of use of those terms changed over time, I was not able to identify any other large-scale trends from the model that might support an art historical reading of titling practice. For instance, the most important words in some topics split into groups in common use at different times, and so to look at the topic proportions in each decade as a measure of the level of use of all the important words would be misleading. In addition, some words that would have been used for similar purposes such as numerals were split across several topics. For these reasons I have not looked to use these results.

I also gathered together titles by the year in which the works they named were created, and topic modelled the resulting 119 documents. Finally, I randomised the order of titles in each decade and output from my dataset a series of documents each containing 500 words, extended if necessary to avoid breaks in the final title. In total, there were over 300 documents. The results of topic modelling these two sets of documents were typically similar to those obtained by modelling each decade as a single text. For instance, models often had the same three dominant topics that exhibited a similar cyclic pattern of change. I decided against using these results as they did not appear to be producing new and useful information. Using them would also have meant presenting consolidated information such as average topic proportions when looking at the trends by decade. It is much more perspicuous to present the results from modelling each decade as a unit. As importantly, it would have precluded me from making quantitative comparisons and interpretations using the entropy of the topic model distributions for each decade as input. The average of all the topic distributions for all the 500-word documents for a decade or all titles used in a particular year are not themselves topics distribution and so could not have been used in that way.

⁸ For a discussion of this issue see Mark Algee-Hewitt, Ryan Heuser, and Franco Moretti, 'Scale, Theme, and Narrative Form', *Stanford Literary Lab Pamphlets*, 10, (2015).

Once the topic modelling application has been 'trained' against some documents with a model developed meeting the needs of the researcher, new documents can be matched against that model. Taking the topics as fixed, the application models each of the new documents as a probability distribution over the topics of the original model determining the weights of each distribution to give the best fit. I used the MALLET application as an 'inferential' tool in this way to look at the titles in my dataset split by geography and by gender.

C.3 Parts-of-Speech Tagging

The second predictive natural language processing technique I used in my reading of titles in modern and contemporary art was parts-of-speech tagging.⁹ Using the same set of documents as for the topic modelling, tagging allowed me to investigate the use of the syntactic components of titles I excluded from the topic model through my choice of stop words, including the definite and indefinite articles, and punctuation marks. Their use may be of art historical significance. For instance, in Chapter 3 I argued that the shift away from the use of the indefinite article at the start of catalogue entries could be interpreted as one measure of the emergence of titling. It also provided a different and complementary perspective to the topic model on the use of other word types such as nouns, adjectives, and verbs.

As words in the English language can have more than one part of speech, for instance 'book' is both a noun and a verb, and as typically up to two-thirds of the words in a text are ambiguous in this way, parts-of-speech tagging is a disambiguation task. It involves choosing the most likely part of speech to tag a 'syntactic token' of a text in its context.¹⁰ These tokens include the words in the text and other syntactic elements such as numerals, symbols, and punctuation marks.

There are numerous approaches taken to parts of speech tagging and the one I used is the application developed by the Stanford University Natural Language Processing Group, which can be downloaded from the Group's website.¹¹ The statistical model assumed by this tagger is that, given a token to be classified, the probability it is of a certain class is related to various features of the token itself and of the sequence of tokens in which it appears. For tokens such

⁹ For a technical discussion of parts-of-speech tagging, see Daniel Jurafsky and James H. Martin, *Speech and Language Processing*, (London: Pearson, 2008), pp. 157 - 246.

¹⁰ In the literature instances of the different parts of speech are often referred to as 'tokens', however these are not the same as the 'tokens' as defined in my topic modelling and to avoid confusion I refer to them as 'syntactic tokens'.

¹¹ The Stanford University Natural Language Processing Group parts-of-speech tagger can be found at <https://nlp.stanford.edu/software/tagger.shtml>.

as punctuation marks and numerals the tagger effectively tags each occurrence as they are unambiguous. For other tokens, a feature of a word may include its spelling and a feature of the context may be that it appears after a word of a certain class. The weights to give to these features in the model are taken to be those giving the best fit to a pre-tagged 'training set'. As the performance of the model is judged against that standard, parts-of-speech tagging is an example of supervised machine learning.

The version of the Stanford parser I used takes as its weights those derived from a training set including the Wall Street Journal online corpus and other parsing data. It uses the Penn Treebank tagging set of 45 classes of syntactic token.¹² In my use I grouped these together into the main categories of noun, verb, adjective, adverb, pronoun, and determiners (such as the articles), as well as numerals and punctuation marks. Given a new text to classify, the tagger takes the weights from the training set, and, looking at each syntactic tokens, calculates the probability that it is of each part of speech. For tokens such as numerals and punctuation marks the tagger effectively tags each occurrence through assigning a probability of one. For other tokens the algorithm tags it with the part of speech having the highest predicted probability. The tagger does not report the estimated probability of each assignment, neither does it report the estimated probabilities for other parts of speech. Unlike the MALLET topic modelling tool, the Stanford parts-of-speech tagger utilises a deterministic algorithm and will always assign the same part of speech to each syntactic token in a document each time it is run with weights derived from the same training set. The Stanford tagger has a reported accuracy of 97% in assigning parts of speech for English. A lower level of performance with titles is to be expected as titles are often short and so provide less context to the tagger than other texts. However, in my reading I grouped together all kinds of adjectives and nouns as 'objective' parts of speech, and so this measure is unaffected by mis-taggings such as singular nouns as plurals, nouns as proper nouns, and nouns as adjectives. Similarly, the 'subjective' parts of speech include verbs of all kinds, and so the measure is unaffected by mis-taggings such as past tense verbs as past participles. Some mis-taggings such as nouns as verbs and adverbs as adjectives will have affected the subjective and objective measures. To look at how well the tagger was identifying the subjective and objective parts of speech I compared the results of using the tagger with my subjective assignments for 500 randomly selected titles. Overall it was 98% correct.

¹² Jurafsky and Martin, (2008), pp. 160 - 161.

In interpreting the results of the use of a parts-of-speech tagger it is important to understand it works on the assumption the patterns of use of different parts of speech in the training set are largely reproduced in the language used in the texts it is classifying. This may not be so if the training set and the texts to be modelled are widely separated culturally or temporally. The tagger with weights from that training set would not be an appropriate tool to use in such circumstances. However, this is unlikely in my case.

REFERENCES

- Acemoglu, D., and Robinson, J., *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*, (New York: Crown, 2012).
- Adams, S. R., "Noising things abroad": Art, Commodity and Commerce in Post-Revolutionary Paris', *Nineteenth-Century Art Worldwide*, 12, (2013). <http://www.nineteenth-centuryartworldwide.org/autumn13/adams-on-art-commodity-and-commerce-in-post-revolutionary-paris>. Last accessed: July 19, 2021.
- Agnello, R. J., and Pierce, R. K., 'Financial Returns, Price Determinants, and Genre Effects in American Art Investment', *Journal of Cultural Economics*, 20/4, (1996), pp. 359 - 383.
- Agresti, A., *Categorical Data Analysis*, (Hoboken: Wiley, 2013).
- Albert Oehlen, 'Born to be Late', in *Selections from the Collection of Pierre Huber*, (Christie's, 2007). Exhibition Catalogue. <https://www.christies.com/en/lot/lot-4871326>. Last accessed: July 20, 2021.
- Albert Oehlen, 'Treppe', in *Contemporary Art Day Auction*, (Sotheby's, 2019). Exhibition Catalogue. <https://www.sothebys.com/en/auctions/ecatalogue/2019/contemporary-art-day-n10150/lot.449.html>. Last accessed: July 20, 2021.
- Algee-Hewitt, M., Heuser, R., and Moretti, F., 'Scale, Theme, and Narrative Form', *Stanford Literary Lab Pamphlets*, 10, (2015).
- Allison, S., Gemma, M., Hueser, R., Moretti, F., Tevel, A., and Yamboliev, I., 'Style at the Scale of the Sentence', *Stanford Literary Lab Pamphlets*, 5, (2013). <https://litlab.stanford.edu/pamphlets/>. Last accessed July 20, 2021.
- Ambrosini, L., 'Genre Painting under The Restoration and July Monarchy: The Critic Confronts Popular Art', *Gazette des Beaux-Arts*, 125, (1995), pp. 41 - 52.
- Anderson, R. C., 'Paintings as an Investment', *Economic Inquiry*, 12/1, (1974), pp. 13 - 26.
- Anfam, D., *Mark Rothko: The Works on Canvas*, (New Haven: Yale University Press, 1998).
- Ashenfelter, O., and Graddy, K., 'Auctions and the Price of Art', *Journal of Economic Literature*, 41/3, (2003), pp. 763 - 787.

Atkins, G., *Asger Jorn: Catalogue Raisonné*, Vols 1 to 4, (London: Lund Humphries, 1968 to 1980).

Aubry, M., Kräussel, R., Manso, G., and Spaenjers, C., 'Machines and Masterpieces: Predicting Prices in the Auction Market', HEC Paris Research Paper No. FIN-2019-1332, March, (2019). <http://dx.doi.org/10.2139/ssrn.3347175>. Last accessed: July 20, 2021.

Baetens, J. D., 'Vanguard Economics, Rearguard Art: Gustave Coûteaux and the Modernist Myth of the Dealer-Critic System', *Oxford Art Journal*, 33/1, (2010), pp. 27 - 41.

Bailey, J., 'Can Machine Learning Predict the Price of Art at Auction?', *Harvard Data Science Review*, 2/2, (2020).

Bann, S. 'Questions of Genre in French Painting', *New Literary History*, 34/3, (2003), pp. 501 - 511.

Bann, S., 'The Mythical Conception is the Name: Titles and Names in Modern and Postmodern Painting', *Word and Image*, 1/2, (1985), pp. 176 - 190.

Baris Kandemir, J. Z., and Li, J., 'Computerized Analysis of Paintings', in Brown, K., (ed.), *Routledge Companion to Digital Humanities and Art History*, (London: Routledge, 2020). pp. 299 - 313.

Barthes, R., Bastian, H., and Armstrong, T., *Cy Twombly, Paintings and Drawings 1954 - 1977*, (New York: The Whitney Museum of American Art, 1979).

Bastian, H., (ed.), *Cy Twombly: Catalogue Raisonné of the Paintings*, Vols I to VII, (Munich: Schirmer/Mosel, 1997 to 2018).

Baudelaire, C., *Ouvres Complet de Charles Baudelaire, II Curiosités Esthétiques*, (Paris: Michel Lévy Frères, 1868).

Beggs, A., and Graddy, K., 'Anchoring Effects: Evidence from Art Auctions', *The American Economic Review*, 99/3, (2009).

Bell, C., *Art*, (Delhi: Lector House, 2020).

Besa Camprubi, J., *Nouveaux actes semiotiques N° 82: Les fonctions du titre*, (Limoges: Presses Universitaires de Limoges, 2002).

- Bishop, C., 'Against Digital Art History', *International Journal for Digital Art History*, Issue 3, (2018). <https://doi.org/10.11588/dah.2018.3.49915>. Last accessed: July 19, 2021.
- Blei, D. M., Ng, A., and Jordan, M., 'Latent Dirichlet Allocation', *Journal of Machine Learning Research*, 3, (2003), pp. 993 - 1,022.
- Blei, D. M., 'Probabilistic Topic Models', *Communications of the ACM*, 55/4, (2012).
- Boime, A., 'The Salon des Refusés and the Evolution of Modern Art', *Art Quarterly*, 32, (Winter 1969), pp. 411 - 426.
- Boime, A., *The Academy and French Painting in the Nineteenth Century*, (New Haven and London: Yale University Press, 1971).
- Boime, A., *Art in an Age of Counter-Revolution, 1815 - 1848*, (Chicago: Chicago University Press, 2004).
- Bois, Y-A., Buchloh, B., Foster, H., and Krauss, R., *Art Since 1900: Modernism, Anti-Modernism, Postmodernism*, (New York and London: Thames & Hudson, 2004).
- Boyd-Graber, J., Hu, Y., and Mimno, D., 'Applications of Topic Models', *Foundations and Trends in Information Retrieval*, XX/XX, (2017), pp. 1 - 154.
- Braden, L. E. A., and Teekens, T., 'Reputation, Network Status and the Art Market', *Arts*, 8/3, (2019), pp. 81 - 95. <https://doi.org/10.3390/arts8030077>. Last accessed: July 19, 2021.
- Briggs, J., 'Condemned to Sparkle: The Reception, Presentation and Production of Léon Bonvin's Floral Still Lifes', *Oxford Art Journal*, 38/2, (2015), pp. 247 - 262.
- Brinson, K. and Hudson, S., (eds.), *Christopher Wool*, (New York: Guggenheim Museum Publications, 2013).
- Burchett-Lere, D., (ed.), *Sam Francis: Catalogue Raisonné of Canvas and Panel Paintings, 1946 - 1994*, (Berkeley: University of California Press, 2011).
- Butler, C., and Schwartz, A., (eds), *Modern Women: Women Artists at the Museum of Modern Art*, (New York: MOMA, 2012).
- Calder, A., *Calder: An Autobiography with Pictures* (New York: Pantheon Books, 1966).

Carrier, D., *Principles of Art History Writing*, (University Park: The Pennsylvania State University Press, 1994).

Catalogue de la 39^e Exposition, (Paris: Société des Artistes Indépendants, 1928). <https://gallica.bnf.fr/ark:/12148/bpt6k889206x?rk=85837;2>. Last accessed: July 21, 2021.

Catalogue de Peinture, Dessin, Sculpture, Gravure, Architecture et Arts Décoratifs, (Paris: Société du Salon d'Automne, 1904). <https://archive.org/details/cataloguedesouvr1904salo/page/n9>. Last accessed: July 21, 2021.

Chagnon-Burke, V., 'Looking at and Buying Contemporary Art in Mid-Nineteenth-Century Paris', *Nineteenth-Century Art Worldwide*, 11, (2012). <http://www.nineteenth-centuryartworldwide.org/summer12/veronique-chagnon-burke-looking-at-and-buying-contemporary-art-in-mid-nineteenth-century-paris>. Last accessed: July 19, 2021.

Chen, R., Lui, H., and Altmann, G., 'Entropy in Different Text Types', *Digital Scholarship in the Humanities*, 32/3, (2017), pp. 528 - 542. <https://doi.org/10.1093/lc/fqw008>. Last accessed: July 20, 2021.

Coddington, J., Elton, J., Rockmore, D., and Wang, Y., 'Multi-fractal analysis and authentication of Jackson Pollock paintings', *Proc. SPIE*, 6810, (2008), pp. 1 - 12.

Connell, R. W., and Pearse, R., *Gender: In World Perspective*, (London: Wiley and Sons, 2014).

Cover, T. M., and Thomas, J. A., *Elements of Information Theory*, (New York: John Wiley & Sons Inc., 1991).

Crowley, E. J., and Zisserman, A., 'The Art of Detection', in Hua, G., and Dégou, H., (eds.), *ECCV Workshops, Part 1*, (Basel: Springer, 2016), pp. 721 - 737. https://doi.org/10.1007/978-3-319-46604-0_50. Last accessed: July 19, 2021.

DaCosta Kaufmann, T., Dossin, C., and Joyeux-Prunel, B., *Circulations in the Global History of Art*, (London: Ashgate, 2017).

Danto, A., 'The Artworld', *The Journal of Philosophy*, 61/19, (1964), pp. 571 - 584.

de Font-Réaulx, D. *Painting and Photography: 1839 - 1914*, (Paris: Flammarion, 2012).

de la Barre, M., Ducclo, S. and Ginsburgh, V., 'Returns of Impressionist, Modern and Contemporary European Paintings, 1962 - 1991', *Annales d'Économie et de Statistique*, 35, (Jul -Sep. 1994), pp. 143 - 181.

de la Rosa, J., and Suárez, J. L., 'A Quantitative Approach to Beauty. Perceived Attractiveness of Human Faces in World Painting', *International Journal for Digital Art History*, Issue 1, (2015). <https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/21640>. Last accessed: July 19, 2021.

de Valenciennes, P-H., *Éléments de Perspective Pratique, à l'usage des Artistes, suivis de Réflexions et Conseils à un Elève sur la Peinture, et particulièrement sur le genre de Paysage*, (Paris: Desenne, Duprat, 1800).

Dickie, G., *Art and the Aesthetic: An Institutional Analysis*, (New York: Cornell University Press, 1974).

Dossin, C. and Joyeux-Prunel, B., 'The German Century: How a Geopolitical Approach could Transform the History of Modernism', in Kaufmann, T. D., Dossin, C., and Joyeux-Prunel, B. (eds.), *Circulations in the Global History of Art*, (New York: Ashgate, 2017), pp. 183 - 201.

Doy, G. *Women & Visual Culture in 19th Century France: 1800 - 1852*, (London and New York: Leicester University Press, 1998).

Drucker, J., 'A Conversation on Digital Art History', in. Gold, M. K., and Klein, L. F., (eds.), *Debates in the Digital Humanities 2019*, (Minneapolis: University of Minnesota Press, 2019), pp. 321 - 334.

Drucker, J., 'Is There a "Digital" Art History?', *Visual Resources: An International Journal of Documentation*, 29:1-2, (2013), pp. 5 - 13. [10.1080/01973762.2013.761106](https://doi.org/10.1080/01973762.2013.761106). Last accessed July 19, 2021.

Dunn, S., and Woolford, K., 'Reconfiguring experimental archaeology using 3D reconstruction', *Proceedings of the 2012 Conference on Electronic Visualization and the Arts*, (New York: Springer, 2013), pp. 277 - 291.

Elgammal, A., Mazzone, M., Liu, B., Kim, D., and Elhoseiny, M., 'The Shape of Art History in the Eyes of the Machine', 32nd AAAI Conference on Artificial Intelligence, New Orleans, USA, February 2 - 7, 2018. [arXiv:1801.07729](https://arxiv.org/abs/1801.07729). Last accessed: July 19, 2021.

Elsner, J., *Style*, in Nelson, R. S., and Schiff, R., (eds.), *Critical Terms for Art History*, (Chicago: University of Chicago Press, 2003), pp. 98 - 109.

Feinstein, C. H. and Thomas, M., *Making History Count: a primer in quantitative methods for historians*, (Cambridge: Cambridge University Press, 2009).

Fletcher, P., and Heimrich, A., 'Local/Global: Mapping Nineteenth-Century London's Art Market', *Nineteenth-Century Art Worldwide*, 11/3, (2012). <https://www.19thc-artworldwide.org/autumn12/fletcher-helmreich-mapping-the-london-art-market>. Last accessed: July 19, 2021.

Forest, D., Truong, V., and Lemay, Y., 'De quoi est-il question dans le discours en art contemporain? La fouille de textes appliquée à l'art contemporain dans les centres d'artistes.', *Digital Humanities 2017*, Montréal, Canada, August 8 - 11, 2017. <https://dh2017.adho.org/program/abstracts/>. Last accessed: July 19, 2021.

Freund, A. *Portraiture and Politics in Revolutionary France*, (University Park: The Pennsylvania State University Press, 2014).

Fritsch, L., (ed.), *A. R. Penck: I Think in Pictures*, (Oxford: Ashmolean, 2019).

Galenson, D. W. and Jensen, R., 'Careers and Canvases: The Rise of the Market for Modern Art in the Nineteenth Century', in *Current Issues in Nineteenth Century Art, Van Gogh Studies*, 1, (Waanders: Zwolle, 2007).

Galenson, D., 'The Careers of Modern Artists: Evidence from Auctions of Contemporary Paintings', NBER Working Paper 6331, December, (1997). <http://www.nber.org/papers/w6331>. Last accessed: July 20, 2021.

Gallo, A., 'A Refresher on Statistical Significance', in *HBR Guide to Data Analytics for Managers*, (Cambridge: Harvard Business Review, 2018), pp. 182 - 185.

Garcia-Zorita, C., and Pacios, A. R., 'Topic modelling characterization of Mudejar art based on document titles', *Digital Scholarship in the Humanities*, 33/3, (2018), pp. 529 - 539. DOI: <https://doi.org/10.1093/lc/fqx055>. Last accessed: July 19, 2021.

Gardner-Huggett, J., 'Extrapolating Influence: The Challenges of Mapping the History of ARC and Artemisia Galleries, Chicago (1980 - 1985)', *Historical Geography*, 45, (2017), pp. 37 - 65.

Genette, G., Crampé, B., (trans.), 'Structure and Function of the Title in Literature', *Critical Inquiry*, 14/4, (1988), pp. 692 - 720.

Genette, G., Lewin, J. E., (trans.), *Paratexts: Thresholds of Interpretation*, (Cambridge: Cambridge University Press, 1997).

Getsy, D. J., *Abstract Bodies: Sixties Sculpture in the Expanded Field of Gender*, (New Haven and London: Yale University Press, 2015).

Godfery, M. and Serota, N., (eds.), *Gerhard Richter: Panorama*, (London: Tate Publishing, 2012).

Godfrey M., Tate Modern Panorama Symposium, 4 April 2012, <https://www.youtube.com/playlist?list=PL906E6634732C1824>. Last accessed: July 20, 2021.

Goldstone, A. and Underwood, T. 'The Quiet Transformation of Literary Studies: What Can Topic Models of PMLA Teach Us About the History of Literary Scholarship?', *Journal of Digital Humanities*, 2/1, 2012. <https://dx.doi.org/10.1353/nlh.2014.002>. Last accessed: July 19, 2021.

Gombrich, E., 'Image and Word in Twentieth Century Art', *Word and Image*, 1/3, (1985), pp. 213 - 241.

Grab, A. *Napoleon and the Transformation of Europe*, (New York: Palgrave, 2003).

Green, N. *The spectacle of NATURE. Landscape and bourgeois culture in nineteenth-century France*, (Manchester: Manchester University Press, 1990).

Green, N., 'Circuits of Production, Circuits of Consumption: The Case of Mid-Nineteenth-Century French Art Dealing', *Art Journal*, 48/1, (1989).

Green, N., 'Dealing in Temperaments: Economic Transformation of the Artistic Field in France during the Second Half of the Nineteenth Century', *Art History*, 10/1, (1987).

Greenacre, M., *Correspondence Analysis in Practice*, (2nd ed.), (London and New York: Chapman and Hall, 2007).

Greenberg, C., O'Brian, J., (ed.), *The Collected Essays and Criticism, Vols 1 to 4*, (Chicago: Chicago University Press, 1988).

Greet, M., *Transatlantic Encounters: Latin American Artists in Paris between the Wars*, (New Haven and London: Yale University Press, 2018).

Griffiths, H., 'The Academy and the Salon Jury, 1831 - 1848', in Kearns and Mill, (eds.), (2015), pp. 64 - 86.

Griffiths, H., *The Jury of the Paris Fine Art Salon, 1831 to 1852*, (doctoral thesis, University of Exeter, 2013).

Guyot de Fère, M., *Statistique des Beaux-Arts en France*, (Paris: Doney-Dupré, 1835).

Halbreich, K., (ed.), *Alibis: Sigmar Polke 1962-2010*, (London: Tate Publishing, 2014).

Halliday, T., *Facing the Public: Portraiture in the Aftermath of the French Revolution*, (Manchester: Manchester University Press, 2000).

Hauptman, W., 'Juries, Protests and Counter-Exhibitions', *The Art Bulletin*, 67/1, (1985).

Helen Higgs and John Forster, 'The Auction Market for Artworks and their Physical Dimensions including the Golden Ratio: Australia - 1986 - 2009', SSRN Research Paper, June, (2011). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1867836. Last accessed: July 20, 2021.

Higgs, H., and Worthington, A., 'Financial Returns and Price Determinants in the Australian Art Market: 1973 - 2003', *The Economic Record*, 81/253, (2005), pp. 113 - 123.

Hock, B., and Owen, J., *A Reader in East-Central European Modernism, 1918 - 1956*, (London: Courtauld Books Online, 2019). <https://courtauld.ac.uk/research/courtauld-books-online/a-reader-in-east-central-european-modernism-1918-1956>. Last accessed: July 20, 2021.

Hodgson, D. J., 'Age-price profiles for Canadian painters at auction', *Journal of Cultural Economics*, 35, (2011), pp. 287 - 308.

Hoek, L. H., *Titres, toiles et critique d'art, Déterminants institutionnels du discours sur l'art aux dix-neuvième siècle en France*, (Leiden: Brill, 2001).

Hopkins, D., *After Modern Art: 1945 - 2017*, (Oxford: Oxford University Press, 2018).

Horowitz, N., *Art of the Deal: Contemporary Art in a Global Financial Market*, (Princeton and Oxford: Princeton University Press, 2011).

House, J., 'Authority versus independence: the position of French landscape in the 1870s', in Richard Thompson, (ed.), *Framing France: Essays on the Representation of Landscape in France, 1870 - 1914*, (Manchester: Manchester University Press, 1998), pp. 18 - 34.

Hunt, L., *Politics, Culture and Class in the French Revolution*, (Berkeley: University of California Press, 1984).

Jacobus, M., *Reading Cy Twombly: Poetry in Paint*, (Princeton and Oxford: Princeton University Press, 2019).

Jaskot, P. B., 'Digital Methods and the Historiography of Art', in Kathryn Brown, (ed.), *The Routledge Companion to Digital Humanities and Art History*, (London: Routledge, 2020), pp. 9 - 19.

Jockers, M. L., *Macroanalysis: Digital Methods and Literary History*, (Chicago: University of Illinois Press, 2013).

Jonker, M., 'Gustave Planche or The Romantic Side of Classicism', *Nineteenth-Century Art Worldwide*, 3/1, (2004). <http://www.19thc-artworldwide.org/spring04/82-autumn02/autumn02article/257-gustave-planche-or-the-romantic-side-of-classicism>. Last accessed: July 19, 2021.

Joyeux-Prunel, B., 'Provincializing Paris, the Center-Periphery Narrative of Modern Art in the Light of Quantitative and Transnational Approaches', *ARTL@S Bulletin*, 4/1, (2015), Article 4. <https://docs.lib.purdue.edu/artlas>. Last accessed: July 19, 2021.

Juola, P., 'Using the Google N-Gram Corpus to measure cultural complexity', *Literary and Linguistic Computing*, 28/4, (2013), pp. 668 -675.

Jurafsky, D. and Martin, J. H., *Speech and Language Processing*, (London: Pearson, 2008).

Kalb, P. R., *Art Since 1980: Charting the Contemporary*, (London: Lawrence King, 2013).

Kearns J., and Mill, A., (eds.), *The Paris Fine Art Salon, 1791 - 1881*, (Bern: Peter Lang, 2015).

Kearns, J., 'No Object too Humble? Still Life Painting in French Art Criticism during the Second Empire', in Rigby, B., (ed.), *French Literature, Thought and Culture in the Nineteenth Century*, (London: Macmillan, 1993), pp. 148 - 168.

Khong, W-H, Soon, L-K, Goh, H-N, Haw, S-C, (2018) 'Leveraging Part-of-Speech Tagging for Sentiment Analysis in Short Texts and Regular Texts', in Ichise, R., Lecue, F., Kawamura, T., Zhao, D., Muggleton, S., Kozaki, K., (eds.) *Semantic Technology. JIST 2018. Lecture Notes in Computer Science*, Vol. 11341, (New York, Springer, 2018) and Jockers, (2012), pp. 118 - 153.

Klinke, H., 'The Digital Transformation of Art History', in Kathryn Brown, (ed.), *The Routledge Companion to Digital Humanities and Art History*, (London: Routledge, 2020), pp. 32 - 42.

Kolešnik, L., 'The Transition of New Tendencies from Neo-Avant-Garde Subculture to Institutional Mainstream Culture. An Example of Network Analysis', in Kolešnik, L. and Horvatinčić, S., (eds.), *Modern and Contemporary Artist Networks*, (Zagreb: Institute of Art History, 2018), pp. 84 - 112.

Kurczynski, K., *The Art and Politics of Asger Jorn: The Avant-Garde Won't Give Up*, (London: Ashgate, 2014).

Larry's List, Private Art Museum Report, (Hong Kong: Larry's List, 2015).

Lavezzi, E., 'Typologie et Imaginaire du Paysage dans les *Éléments* de Valenciennes, Hériet de Pies et Watelet', in Grosrichard, A., (ed.), *Dénouement des lumières et invention romantique: actes du colloque, Genève, 24-25 November 2000*, (Geneva: Librairie Droz, 2003).

Lévesque, P-C. and Watelet, C-H., *Dictionnaire des Arts de Peinture, Sculpture et Gravure*, (Paris: Prault, 1792).

Lincoln, M., 'If Paintings were Plants: Measuring Genre-Diversity in Seventeenth Century Dutch Painting and Printmaking', Digital Humanities 2016, Kraków, Poland, 11 - 16 July, 2016. <https://dh2016.adho.org/abstracts/133>. Last accessed: July 19, 2021.

Lincoln, M., 'Social Network Centralization on Print Production in the Low Countries, 1550 - 1750', *International Journal for Digital Art History*, Issue 2, (2016). <https://doi.org/10.11588/dah.2016.2.25337>. Last accessed: July 19, 2021.

Long, J. S., and Ervin, L. H., 'Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model', *The American Statistician*, 54/3, (2000), pp. 217 - 224.

Lord, C., and Meyer, R., *Queer Art and Culture*, (2nd ed.), (London: Phaidon Press, 2019).

Mainardi, P., *Art and Politics of the Second Empire*, (New Haven: Yale University Press, 1987).

Mainardi, P., *The End of the Salon: Art and the State in the Early Third Republic*, (Cambridge: Cambridge University Press, 1993).

Mason, B., 'What is Driving the Soaring Demand for Art Storage?', *Apollo: The International Art Magazine*, 27 June 2017, pp. 67 - 68.

Mayo Roos, J., *Early Impressionism and the French State*, (Cambridge: Cambridge University Press, 1996).

McCoubrey, J. W., 'The Revival of Chardin in French Still-Life Painting, 1850-1870', *The Art Bulletin*, 46/1, (1964), pp. 39 - 51.

McPherson, H., (ed.), *The modern portrait in nineteenth-century France*, (Cambridge: Cambridge University Press, 2001).

McWilliam, N., (ed.), *A Bibliography of Salon Criticism in Paris from the Ancien Regime to the Restoration, 1699 - 1827*, (Cambridge: Cambridge University Press, 1991).

McWilliam, N., (ed.), *A Bibliography of Salon Criticism in Paris from the July Monarchy to the Second Republic, 1831 - 1851*, (Cambridge: Cambridge University Press, 1991).

Mehring, C., 'Richter's Willkür', *artjournal*, Winter, (2012), pp. 21 - 35.

Mills, A., 'Artists at the Salon During the July Monarchy', in Kearns and Mill, (eds.), (2015), pp. 45 - 63.

Monnier, G., *L'art et ses institutions en France: De la Révolution à nos jours*, (Paris: Gallimard, 1994).

Moravec, M., 'Network Analysis and Feminist Artists', *ARTL@S Bulletin*, 6/3, (2017), Article 3. <https://docs.lib.purdue.edu/artlas>. Last accessed: July 19, 2021.

Mullenix, J. W., and Robinet, J., 'Art Expertise and the Processing of Titled Abstract Art', *Perception*, 47(4), (2018), pp. 359 - 378.

Nijboer, H., Brouwer, J., and Jan Bok, M., 'Unthinking Rubens and Rembrandt: Counterfactual Analysis and Digital Art History', Digital Humanities 2019, Utrecht, The Netherlands, 8 - 12 July, 2019. <https://dev.clariah.nl/files/dh2019/boa/0980.html>. Last accessed: July 19, 2021.

Nochlin, L., *Women, Art and Power*, (New York: Harper and Rowe, 1988).

Nochlin, L., 'Joan Mitchell a Rage to Paint', in Livingston, J., (ed.), *The Paintings of Joan Mitchell*, (Berkeley: University of California Press, 2002), p. 58.

O'Brien, E., Nicodemus, E., Chiu, M., Genocchio, B., Coffey, M. K., and Tejada, B., (eds.), *Modern Art in Africa, Asia and Latin America*, (Oxford: Blackwell Publishing, 2013).

Oosterlinck, K., and Radermecker, A-S., "'The Master of ...': creating names for art history and the art market", *Journal of Cultural Economics*, 43/1, (2019), pp. 57 - 95.

Orwicz, M. R., (ed.), *Art Criticism and its Institutions in Nineteenth-Century France*, (Manchester: Manchester University Press, 1994).

Panofsky, E., *Studies in Iconology: Humanistic Themes in the Art of the Renaissance*, (London: Routledge, 1972)

Parker, R. and Pollock, G., *Old Mistresses: Women, Art and Ideology*, (London: Pandora Press, 1981).

Parsons, C. V., and Ward, M., (eds.), *A Bibliography of Salon Criticism in Second Empire Paris*, (Cambridge: Cambridge University Press, 1986).

Piper, A., *enumerations: Data and Literary Study*, (Chicago: University of Chicago Press, 2018).

Quach McCabe, S., 'Intermediaries and the market: Hans Rottenhammer's use of networks in the Copper Painting Market', *Arts*, 8/3, (2019), pp. 75 - 96. <https://doi.org/10.3390/arts8020075>. Last accessed: July 19, 2021.

Rasterhoff, C., Beelen, K., Li, W., and Kisjes, I., 'Measuring Innovation in the Art and Book Market during the Dutch Golden Age', Digital Humanities Benelux, Amsterdam, the Netherlands, 7 - 8 June 2018. <http://2018.dhbenelux.org/programme/detailed-programme/>. Last accessed: July 19, 2021.

Renneboog, L., and Spaenjers, C., 'Buying Beauty: On Prices and Returns in the Art market', *Management Science*, 59/1, (2013), pp. 36 - 53.

Riegl, A., Kain, E., (trans.), *Problems of Style: Foundations for a History of Ornament*, (Princeton: Princeton University Press, 2018).

Rockwell, G. and Sinclair, S., *Hermeneutica: Computer Assisted Interpretation in the Humanities*, (Cambridge: MIT Press, 2016).

Rodriguez Ortega, N., 'Digital Art History: An Examination of Conscience', *Visual Resources: An International Journal of Documentation*, 29:1-2, (2013), pp. 129 - 133. [10.1080/01973762.2013.7611024](https://doi.org/10.1080/01973762.2013.7611024). Last accessed: July 19, 2021.

Rodriguez Ortega, N., 'Digital Art History: The Questions that Need to be Asked', *Visual Resources*, 35, (2019), pp. 6 - 20.

Rodriguez Ortega, N., 'Image Processing and Computer Vision in the Field of Art History', in Kathryn Brown, (ed.), *The Routledge Companion to Digital Humanities and Art History*, (2020). pp. 338 - 357.

Saint-Martin, F., *Semiotics of Visual Language*, (Indianapolis: Indiana University Press, 1990).

Saint-Raymond, L., 'Revisiting Harrison and Cynthia White's Academic vs. Dealer-Critic System', *Arts*, 8/3, (2019), pp. 96 - 113. <https://doi.org/10.3390/arts8030096>. Last accessed: July 19, 2021.

Saleh, B., Abe, K., Singh Arora, R., and Elgammal, A., 'Toward automated discovery of artistic influence', *Multimedia Tools and Applications*, 75, (2014), pp. 3565 - 3591.

Salway, A., and Baker, J., 'Investigating Curatorial Voice with Corpus Linguistic Techniques: the case of Dorothy George and applications in museological practice', *Museum and Society*, 18(2), (July 2020), pp. 151 - 170.

Schich, M., Song, C., Ang, Y-Y., Mirsky, A., Martino, M., Barabási, A-L., and Helbing, D., 'A network framework for cultural history', *Science*, 345, (2014), pp. 558 - 562. <https://www.youtube.com/watch?v=4glhRkCcD4U>. Last accessed: July 19, 2021.

Schmeuli, G., 'To Explain or to Predict', *Statistical Sciences*, 25/3, (2010), pp. 289 - 310.

Schmidt, B. M., 'Words Alone: Dismantling Topic Models in the Humanities', *Journal of Digital Humanities*, 2/1, 2012. <http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt/>. Last accessed: July 19, 2021.

Schmidt, B. M., 'Do Digital Humanists Need to Understand Algorithms?', in Gold, M. K., and Klein, L. F., *Debates in the Digital Humanities*, (Minneapolis: University of Minnesota Press, 2016), pp. 546 - 555.

Scott Long, J., and Ervin, L. H., 'Using Heteroscedasticity Consistent Standard Errors in the Linear Regression model', *The American Statistician*, 54/3, (2000), pp. 217 - 224.

Seave Greenwald, D., 'Modernization and rural imagery at the Paris Salon', *The Economic History Review*, 72/2, (2019), pp. 531 - 567.

Seave Greenwald, D., *Painting by Numbers: Data-Driven Histories of Nineteenth-Century Art*, (Princeton: Princeton University Press, 2021).

Seber, G. A., *Statistical Models for Proportions and Probabilities*, (London: Springer, 2013).

Shaffer, L. K., 'When the Alpha is the Omega: P-Values, "Substantial Evidence", and the 0.05 Standard at the FDA', *Food Drug Law J*, 72(4), (2017), pp. 595 - 635.

Sheather, S. J., *A Modern Approach to Regression with R*, (New York: Springer, 2010).

Sloane, J. C., *French Painting: Artists, Critics and Traditions from 1848 to 1870*, (Princeton: Princeton University Press, 1951).

Smeets, J., Scholtes, J., Rasterhoff, C., and Schavemaker, M., 'SMTP: Stedelijk Museum Text Mining Project', Digital Humanities 2016, Kraków, Poland, 11 - 16 July, 2016. <https://dh2016.adho.org/abstracts/270>. Last accessed: July 19, 2021.

Smith, T., *Contemporary Art: World Currents*, (London: Laurence King, 2011).

Smith, T., *What is Contemporary Art?*, (Chicago: Chicago University Press, 2009).

Société des "Artistes Indépendants", Catalogue de la 39^e Expositions, (Paris: n.p. 1928). <https://gallica.bnf.fr/ark:/12148/bpt6k889206x?rk=85837;2>. Last accessed: July 20, 2021.

Société du Salon d'Automne, *Catalogue de Peinture, Dessin, Sculpture, Gravure Architecture et Arts Decoratifs*, (Paris, 1904).

<https://archive.org/details/cataloguedesouvr1904salo/page/n10>. Last accessed: July 20, 2021.

Sopranzetti, B., 'Hedonic Regression Models' in Cheng-Few Lee and John C. Lee, (eds.), *Handbook of Financial Econometrics and Statistics*, (New York: Springer, 2014), pp. 2,119 - 2,134.

Stamatatos, E., Kokkinakis, G., and Fakotakis, N., 'Automatic text categorization in terms of genre and author', *Computational Linguistics*, 26, (2000), pp. 471 - 495.

Storr, R., (ed.), *Gerhard Richter: Forty Years of Painting*, (New York: The Museum of Modern Art, 2002).

Stranahan, C. H., *A History of French Painting*, (London: Sampson, Low, Marston, Searle, and Rivington, 1889).

Strezoski, G., and Worring, M., 'OmniArt: A Large-scale Artistic benchmark', *ACM Transactions on Multimedia, Computing, Communications and Applications*, 14/4, (2018), Article 88.

Strucken, M., and Cartwright, L., *Practices of looking: an introduction to visual culture*, (Oxford: Oxford University Press, 2000).

Summer, D., *Style*, in Donald Preziosi, (ed.), *The Art of Art History: A Critical Anthology*, (Oxford: Oxford University Press, 2009), pp. 144 - 150.

Talbot, W. S. and Weisberg, G. R., *Chardin and the Still-Life Tradition in France*, (Cleveland: Cleveland Art Museum, 1979).

The Contemporary Art Market Report, 2019. (Artpprice, 2019).
<https://www.artprice.com/artmarketinsight/reports>. Last accessed: July 19, 2021.

Thompson, R. (ed.), *Framing France; Essays on the Representation of Landscape in France, 1870 - 1914*, (Manchester: Manchester University Press, 1998).

Tintertow, G. and Conisbee, P., (eds.), *Portraits by Ingres: Images of an Epoch*, (New Haven: Yale University Press, 2000).

Underwood, T., *Digital Horizons, Digital Evidence and Literary Change*, (Chicago: Chicago University Press, 2019).

Underwood, T., Bamman, D., and Lee, S., 'The Transformation of Gender in English Language Fiction', *Journal of Cultural Analytics*, February 13, (2018). <http://culturalanalytics.org/2018/02/the-transformation-of-gender-in-english-language-fiction/>.

Last accessed July 19, 2021.

Ursprung, H., W. and Wiermann, C., 'Reputation, price, and death: an empirical analysis of art price formation', CESifo Working Paper, No. 2237, (2008).

Vasari, G., *The Lives of the Artists*, Julia Bondarella and Peter Bondarella, (trans.), (Oxford: Oxford University Press, 1991).

Veltius, O. *Talking Prices: Symbolic Meanings of Prices on the Market for Contemporary Art*, (Princeton and Oxford: Princeton University Press, 2005).

Volikov, R., 'Art Auction prices: Home Bias, Familiarity and Patriotism', SSRN Research paper, August, (2015). <http://dx.doi.org/10.2139/ssrn.2686527>. Last accessed: July 20, 2021.

Welchman, J. C., *Invisible Colors: A Visual History of Titles*, (New Haven and London: Yale University Press, 1997).

White, H. C. and White, C. A., *Canvases and Careers: Institutional Change in the French Painting World*, (New York: J. Wiley and Sons, 1965).

Whiteley, L., 'Art et commerce d'art en France avant l'époque impressioniste', *Romantisme*, 40, (1983), pp. 65 - 76.

Williams, G., 'Jokes Interrupted: Martin Kippenberger's Receding Punch Line', in Kristoff, D., and Morgan, J., (eds.), *Martin Kippenberger*, (London: Tate Publishing, 2006).

Wölfflin, H., Blower, J., (trans.), *Principles of Art History: The Problem of the Development of Style in Early Modern Art*, (Los Angeles: Getty Research Institute Press, 2015).

Wrigley, R., *The Origins of French Art Criticism from the Ancien Regime to the Restoration*, (Oxford: Oxford University Press, 1993).

Yeazell, R. B., *Picture Titles: How and Why Western Paintings Acquired Their Names*, (Princeton: Princeton University Press, 2015).

Yeldham, C. *Women Artists in Nineteenth-Century France and England*, 2 Vols., (London: Garland, 1984).

Yoshitake, M., and Govan, M., *Yoshimoto Nara*, (London: Prestel, 2020).

Zorich, D., M., 'Digital Art History: Community Assessment', *Visual Resources: An International Journal of Documentation*, 29:1-2, (2013), pp. 14 -21.
[10.1080/01973762.2013.761108](https://doi.org/10.1080/01973762.2013.761108). Last accessed: July 19, 2021.

Internet Resources

Alliance of Digital Humanities Organisations: <https://www.adho.org/>. Last accessed: July 19, 2021.

Apollo Magazine: <https://www.apollo-magazine.com/what-is-driving-the-soaring-demand-for-art-storage/>. Last accessed: July 20, 2021.

Artprice.com: <https://www.artprice.com>. Last accessed: July 20, 2021.

British and Irish Furniture Makers Online: <https://bifmo.history.ac.uk/>. Last accessed: July 19, 2021.

British Royal Academy Exhibition Catalogues: <https://www.royalacademy.org.uk/art-artists/search/exhibition-catalogues>. Last accessed: July 20, 2021.

Database of Salon Artists: <http://humanities-research.exeter.ac.uk/salonartists/about>. Last accessed: July 19, 2021.

Digital Humanities Congress: <https://www.dhi.ac.uk/dhc2022/>. Last accessed: October 20, 2021.

Digital Humanities Quarterly: <http://www.digitalhumanities.org/dhq/>. Last accessed: October 19, 2021.

Digital Scholarship in the Humanities: <https://www.academic.oup.com/dsh>. Last accessed: July 19, 2021.

Diogenes: <https://d.iogen.es/d/index.htm>. Last accessed: July 19, 2021.

DisplayR: <https://www.displayr.com/interpret-correspondence-analysis-plots-probably-isnt-way-think/>. Last accessed: July 20, 2021.

Gerhard Richter: <http://www.gerhardrichter.com/>. Last accessed: July 19, 2021.

Getty Research Institute: <https://www.getty.edu/research/>. Last accessed: July 19, 2021.

Google Translate: <https://translate.google.co.uk/>. Last accessed: 20 July, 2021.

Internet Archive: <https://www.archive.org>. Last accessed: July 19, 2021.

Journal of Digital Humanities: <http://journalofdigitalhumanities.org/>. Last accessed: October 20, 2021.

MALLET: MACHine Learning for Language Toolkit: <https://mimno.github.io/Mallet/index>. Last accessed: July 20, 2021.

Mapping Titian: <http://www.mappingtitian.org/about>. Last accessed: July 19, 2021.

Mathematica: <http://www.mathematica-journal.com/2010/09/an-introduction-to-correspondence-analysis/>. Last accessed: July 20, 2021.

Mike Bowman, Writing by Numbers datasets: <https://github.com/MikeBowman-WritingbyNumbers>. Last accessed: July 20, 2021.

Niemanlab: <http://www.niemanlab.org/2011/10/word-clouds-considered-harmful/>. Last accessed: July 19, 2021.

Poetics: <https://www.sciencedirect.com/journal/poetics/vol/41/issue/6>. Last accessed: July 20, 2021.

Proquest: <https://www.proquest.com/>. Last accessed: July 19, 2021.

Python: www.python.org. Last accessed: July 19, 2021.

Republic of Letters: <http://republicofletters.stanford.edu/index.html>. Last accessed: July 19, 2021.

Stanford Literary Lab: <https://litlab.stanford.edu/>. Last accessed: July 19, 2021.

The Artl@s Bulletin: <https://docs.lib.purdue.edu/artlas/>. Last accessed: October 20, 2021.

The Artl@s Project: <https://artlas.huma-num.fr/en/>. Last accessed: October 20, 2021.

The Hathi Trust: <https://www.hathitrust.org/>. Last accessed: July 19, 2021.

The International Journal for Digital Art History: <https://www.dahj.org/articles>. Last accessed: October 20, 2021.

The Stanford Natural Language Processing Group. Stanford Named Entity Recognizer: <https://nlp.stanford.edu/software/CRF-NER.html>. Last accessed: July 20, 2021.

The Stanford Natural Language Processing Group. Stanford Part-of-Speech Tagger: <https://nlp.stanford.edu/software/tagger.shtml>. Last accessed: July 20, 2021.

Visual Resources: <https://www.tandfonline.com/toc/gvir20/current>. Last accessed: October 20, 2021.

Voyant Tools: <https://voyant-tools.org/>. Last accessed: 19 July, 2021.

Criticism of the Paris Salon

All of the resources listed below were last accessed on 12 April 2021.

About, E., *Nos Artistes au Salon de 1857*, (Paris: n. p., 1858).
<https://gallica.bnf.fr/ark:/12148/bpt6k6485867v?rk=42918;4>

Alesson, J., *Les Femmes Artistes au Salon de 1878 et à l'Exposition Universelle*, (Paris: n. p., 1878). <https://gallica.bnf.fr/ark:/12148/bpt6k743037?rk=21459;2>

Anon, *Panard au Salon*, (Paris: n. p., A La Haye, 1781).
<https://gallica.bnf.fr/ark:/12148/btv1b8458027k?rk=21459;2>

Anon, *Loterie pittoresque, pour le Salon de 1873*, (Paris: n. p., 1783).
<https://gallica.bnf.fr/ark:/12148/btv1b8458059j.r=Loterie%20pittoresque%20%20%2C%20pour%20le%20Salon%20de%201783?rk=21459;2>

Anon, *Le Véridique au Sallon*, (Paris: n. p., 1783).
<https://gallica.bnf.fr/ark:/12148/btv1b8458066p.r=Le%20V%C3%A9ridique%20au%20Sallon?rk=21459;2>

Anon, *Le Frondeur, ou Dialogues sur le Sallon*, (Paris: n. p., 1785).
<https://gallica.bnf.fr/ark:/12148/btv1b10523791h?rk=42918;4>

Anon, *Promenades d'un observateur au salon, l'année de 1787*, (Paris: Chez les Marchands de Nouveautés, 1787). <https://gallica.bnf.fr/ark:/12148/btv1b84582775?rk=21459;2>

Anon, *Le Cousin Jacques hors du Sallon, Folie sans consequence, A l'occasion des Tableaux exposés au Louvre en 1787*, (Paris: Chez Royez, 1787).
<https://gallica.bnf.fr/ark:/12148/btv1b8458294p?rk=21459;2>

Anon, *Merlin au Salon en 1787*, (Paris: n. p., 1787).
<https://gallica.bnf.fr/ark:/12148/btv1b84582901?rk=21459;2>

Anon, *Tarare au Sallon de Peinture*, (Paris: Chez les Marchands de Nouveautés, 1787).
<https://gallica.bnf.fr/ark:/12148/btv1b84582812?rk=21459;2>

Anon, *Remarques sur les Ouvrages Exposés au Salon*, (Paris: Chez Knapen Fils, 1789).
<https://gallica.bnf.fr/ark:/12148/bpt6k1655822?rk=21459;2>

Anon, *La Béquille de Voltaire au Salon*, (Paris: Chez les Marchands de Nouveautés, 1791).
<https://gallica.bnf.fr/ark:/12148/btv1b8458345j?rk=21459;2>

Anon, *Lettres Critiques sur les Tableaux du Sallon de 1791*, (Paris: n. p., 1791).
<https://gallica.bnf.fr/ark:/12148/btv1b8458348s?rk=21459;2>

Anon, *Nouvelle Critique Impartiale des Tableaux du Sallon*, (Paris: Chez Bignon, 1791).
<https://gallica.bnf.fr/ark:/12148/btv1b8458350v?rk=21459;2>

Anon, *Sallon de Peinture. 1791.* (Paris: n. p., 1791).
<https://gallica.bnf.fr/ark:/12148/btv1b84583496.r=Sallon%20de%20peinture.%201791?rk=21459;2>

Anon, *Coup-D'Oeil sur Le Sallon du Louvre, de l'an 5^{eme} de la République*, (Paris: n. p., 1796).
<https://gallica.bnf.fr/ark:/12148/btv1b105237618.r=Coup-d%27%27%C5%93il%20sur%20le%20Sallon%20du%20Louvre%2C%20de%20l%27%27an%205eme%20de%20la%20R%C3%A9publique?rk=21459;2>

Anon, *Itineraire Critique du Salon de l'an VI*, (Paris: n. p., 1798).
<https://gallica.bnf.fr/ark:/12148/btv1b10523700b/f1.item.r=Itin%C3%A9raire%20critique,%20du%20Salon%20de%20l'an%20VI>

Anon, *Coup d'Oeil sure le Salon de l'An VIII*, (Paris: Chez Bernard, 1800).
<https://gallica.bnf.fr/ark:/12148/btv1b10524266z?rk=21459;2>

Anon, *Arlequin au Muséum, ou Les Tableaux en Vaudevilles*, (Paris: n. p., 1802).
<https://gallica.bnf.fr/ark:/12148/btv1b10524085k?rk=21459;2>

Anon, *Le Pausanias Français; Salon de 1806*, (Paris: Chez F. Buisson, 1806).
<https://gallica.bnf.fr/ark:/12148/btv1b10545100c?rk=21459;2>

Anon, *Le Flâneur au Salon*, (Paris: Chez M. Aubry, 1806).
<https://gallica.bnf.fr/ark:/12148/btv1b10545083n.r=Le%20fl%C3%A2neur%20au%20salon%20ou%20Mr%20Bon-homme?rk=21459;2>

Anon, 'Salon de 1806', *Mercure de France*, October, (1806).
<https://gallica.bnf.fr/ark:/12148/btv1b10535151k/f1.item.r=Mercure%20de%20France>

Anon, *Observations sur le Salon de l'An 1808*, (Paris: Chez Veuve Geuffier, 1808).
<https://gallica.bnf.fr/ark:/12148/btv1b10545951x?rk=21459;2>

Anon, *La Verité au Salon de 1812*, (Paris: Chez Chassaignan, 1812).
<https://gallica.bnf.fr/ark:/12148/bpt6k62303899?rk=21459;2#>

Anon, *Un tour au Salon or revue critique des tableaux de 1817*, (Paris: Chez Pélicier, 1817).
<https://gallica.bnf.fr/ark:/12148/bpt6k58230917?rk=21459;2>

Anon, *Essai sur le Salon de 1817*, (Paris: Chez Delaunay, Chez Pélicier, 1817).
<https://gallica.bnf.fr/ark:/12148/bpt6k56063982?rk=21459;2>

Anon, 1819, *Letters a David, Sur le Salon de 1819*, (Paris: Chez Pillet Ainé, 1819).
<https://gallica.bnf.fr/ark:/12148/bpt6k62304479?rk=21459;2#>

Anon, *Nicaise, Observateur au Salon de 1822*, (Paris: F.-P. Hardy, 1822).
<https://gallica.bnf.fr/ark:/12148/bpt6k6558151x?rk=21459;2#>

Anon, *Revue Critique des Productions de Peinture, Sculpture, Gravure, Exposées au Salon de 1824*, (Paris: Chez J. G. Dentu, 1824).
<https://gallica.bnf.fr/ark:/12148/bpt6k58235642?rk=21459;2>

Anon, *Salon de mil huit cent vingt-quatre*, (Paris: Pillet Ainé, 1824).
<https://gallica.bnf.fr/ark:/12148/bpt6k6305416q?rk=42918;4>

Anon., *Une Matinée au Salon, ou les peintres de l'école passes en revue*, (Paris: Delaunay, 1824).
<https://gallica.bnf.fr/ark:/12148/bpt6k5440090g.r=une%20matin%C3%A9e%20au%20salon%20ou%20les%20peintures?rk=21459;2>

Anon, 'Beaux-Arts, Exposition du Louvre, 1824', *Journal des débats politique et littéraires*, 1 September 1824, pp. 1.- 6. <https://gallica.bnf.fr/ark:/12148/bpt6k4348817>

Anon, *Examen du Salon de 1827*, (Paris: n. p., 1827).
<https://gallica.bnf.fr/ark:/12148/bpt6k873217r?rk=21459;2#>

Anon, 'Salon de 1833' in *Charivari*, 17 March, (1833), pp. 1 - 2.
<https://gallica.bnf.fr/ark:/12148/bpt6k115118g>

Anon and Chatelain, F., *Prométhéides. Revue du Salon de 1833*, (Paris: n. p., 1833).
<https://gallica.bnf.fr/ark:/12148/bpt6k1183612t?rk=21459;2>

Anon, *Le Musée, Revue du Salon de 1834*, (Paris: A. Ledoux., 1834).
<https://archive.org/details/lemuserevuedus00deca>

Anon, 'Beaux Arts, Salon de 1837', in *Journal des débats politique et littéraires*, 15 March 1837, pp. 1 - 2. <https://gallica.bnf.fr/ark:/12148/bpt6k439429b.item>

Anon, 'Salon de 1837', in *Revue des Deux Mondes*, 15 April 1837, pp. 145 - 176, 148 - 149, 161 - 162, and 225. <https://gallica.bnf.fr/ark:/12148/bpt6k86859x?rk=42918;4>

Anon, 'Salon de 1839', in *Le Magasin Pittoresque*, 1839, pp. 121 - 122, and 145 - 146.
<https://gallica.bnf.fr/ark:/12148/bpt6k314227>

Anon, 'Salon de 1841', in *Écho de la Littérature et des Beaux-Arts en France et à la Étranger*, May 1841, pp. 242 - 244. <https://gallica.bnf.fr/ark:/12148/bpt6k61182353?rk=64378;0>

Anon, 'Le Salon', in *Revue de Paris*, (1867), pp. 111 - 118, and 219 - 228.
<https://gallica.bnf.fr/ark:/12148/cb32857931h/date18670101>

Anon, 'Salon de 1868', in *L'Europe Artiste*, (1868), p. 89.
<https://gallica.bnf.fr/ark:/12148/bpt6k119611g>

Anon, 'Le Salon de 1873', in *Revue des Deux Mondes*, (1873), Vol. 115, pp. 628 - 669.
<https://gallica.bnf.fr/ark:/12148/bpt6k870427?rk=107296;4>

Auvray, L. *Exposition des beaux-arts: Salon de 1867*, (Paris: J. Renouard, 1867).
<https://gallica.bnf.fr/ark:/12148/bpt6k65694163?rk=64378;0>

Aycard, M., and Flocon, F., *Salon de 1824*, (Paris: n. p., 1824).
<https://gallica.bnf.fr/ark:/12148/bpt6k109056n?rk=21459;2#>

Barbier, A., *Salon de 1839*, (Paris: Joubert, 1839).
<https://archive.org/details/salonbarbier00barb>

Baudelaire, C., 'Salon de 1845', and 'Salon de 1846', in *Oeuvres Complètes de Baudelaire. II Curiosités Aesthetic*, (Paris: Michel Levy, 1868), pp. 1 - 75 and pp. 77 - 196.
<https://archive.org/details/in.ernet.dli.2015.340333>

- Besnus, A. *Le Salon de 1875*, (Paris: n. p., 1875).
<https://gallica.bnf.fr/ark:/12148/bpt6k108360c?rk=42918;4>
- Blanc, C., 'Le Salon de 1866', in *Gazette des Beaux Arts*, XX, (1866), pp. 497 - 542.
<https://gallica.bnf.fr/ark:/12148/bpt6k203084g/f5.item>
- Boissin, F. *Études Artistiques, Salon de 1868*, (Paris: n. p., 1868).
<https://gallica.bnf.fr/ark:/12148/bpt6k1090859?rk=42918;4>
- Boutard, M., 'Salon de 1814, N° II', *Journal des débats politiques et littéraires*, (November, 1814), pp. 3 - 4. <https://gallica.bnf.fr/ark:/12148/bpt6k421086k.item>
- Bürger, W., *Salons de T. Thoré, 1845, 1846, 1847, 1848*, (Paris: Jules Renouard, 1870).
https://books.google.co.uk/books?id=451ZAAAAYAAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- Camp, M. du, *Le Salon de 1864*, (Paris: J. de Claye, 1864).
<https://gallica.bnf.fr/ark:/12148/bpt6k64115311?rk=64378;0>
- Cantaloube, A., *Salon de 1864*, (Paris: n. p., 1864).
<https://gallica.bnf.fr/ark:/12148/btv1b52506208r?rk=21459;2>
- Castagnary, J-A., *Salons (1857 - 1870)*, (Paris: n. p., 1892).
<https://gallica.bnf.fr/ark:/12148/bpt6k550068.r=castagnary%20salons?rk=21459;2>
- Challamel, A., *Salon 1840*, (Paris: n. p., 1840).
https://archive.org/details/gri_albumdusalon00chal
- Champfleury, *Oeuvres Posthumes de Champfleury, Salons 1846 - 51*, (Paris: Alphonse Lemerre, 1894). <https://gallica.bnf.fr/ark:/12148/bpt6k5440106w?rk=21459;2>
- Chaumelin, M., *L'Art contemporain*, (Paris: renouard, 1873). <https://bibliotheque-numerique.inha.fr/collection/item/5755-redirection>
- Chauvin, M., *Salon de mil huit vingt-quatre*, (Paris: n. p., 1824).
<https://gallica.bnf.fr/ark:/12148/bpt6k108361r?rk=21459;2>
- Chaussard, P. J-B., *Exposition des ouvrages de Peinture, Sculpture, Architecture, Gravure, dans les salle du Muséum, premier Thermidor an VI*, (Paris: n. p., 1798).
<https://gallica.bnf.fr/ark:/12148/btv1b10524308r/f1.item.r=Beaux-Arts>

- Cherbuliez, V., *Études de littérature et d'art*, (Paris: n. p., 1872).
<https://gallica.bnf.fr/ark:/12148/bpt6k5427422n?rk=493564;4>
- Claretie, J., *L'Art et Les Artistes Français Contemporains, avec un Avant-Propos sur le Salon de a867* (Paris: Charpentier, 1876).
<https://gallica.bnf.fr/ark:/12148/bpt6k65774034?rk=21459;2>
- Clément de Ris, L., 'Le Salon', in *L'Artiste*, February 1851, pp. 3 - 9.
<https://gallica.bnf.fr/ark:/12148/bpt6k2155733?rk=21459;2#>
- Clermont, A. de., *Salon de 1874*, (Paris: La Flèche, 1874).
<https://gallica.bnf.fr/ark:/12148/bpt6k936889v?rk=64378;0>
- Delaborde, H., *Mélanges sur l'art contemporain*, (Paris: n. p., 1866).
<https://gallica.bnf.fr/ark:/12148/bpt6k9399631?rk=21459;2>
- Dauban, C., A., *Le Salon de 1863*, (Paris: Renouard, 1863).
<https://gallica.bnf.fr/ark:/12148/bpt6k5788939b?rk=21459;2>
- Denon, A. M., *Seconde Lettre sur le Salon de 1806*, (Paris: Brasseur Aîné., 1806).
<https://gallica.bnf.fr/ark:/12148/btv1b10535140q?rk=21459;2>
- Destigny, J-F., *Revue Poétique du Salon de 1840*, (Paris: n. p., 1840).
<https://archive.org/details/revuepotiquedes00destgoog/mode/2up>
- Delécluze, M. E. J., *Les beaux-arts dans les deux mondes*, (Paris: Charpentier, 1856).
<https://archive.org/details/lesbeauxartsdan00delgoog>
- Dumas, A., *L'Art et les Artistes Contemporain au Salon de 1859*, (Paris: n. p., 1859).
<https://gallica.bnf.fr/ark:/12148/bpt6k1148602?rk=21459;2>
- Dumesnil, M. H., *Le Salon de 1859*, (Paris: J. Renouard, 1859).
<https://gallica.bnf.fr/ark:/12148/bpt6k6468601s?rk=21459;2>
- Duval, Ch. -L., Eugène Delacroix et Hippolyte Flandrin Parallèle. Le Salon de 1864. (Paris: A. Cochet, 1864). <https://gallica.bnf.fr/ark:/12148/bpt6k65725599?rk=107296;4>
- Etex, A., *Essai d'une Revue Synthétique sur L'Exposition Universelle de 1855*, (Paris: Etex, 1856). <https://gallica.bnf.fr/ark:/12148/bpt6k62095807?rk=21459;2>

Fizelière, A. de la, *Salon de 1850-51*, (Paris: Passard, 1851).
<https://gallica.bnf.fr/ark:/12148/bpt6k62160744?rk=21459;2>

Flaner, W., *Le Salon de 1857*, (Paris: H. Lefèvre, 1857).
<https://gallica.bnf.fr/ark:/12148/bpt6k6217261h.r=salon%20de%201857?rk=107296;4#>

Flécherye, P. de la, *Le Salon de 1875*, (Paris: n. p., 1875).
<https://gallica.bnf.fr/ark:/12148/bpt6k1411151c?rk=21459;2>

Gault de Saint-Germain, P., M., *Choix des productions de l'art les plus remarquables exposées dans le Salon de 1817*, (Paris: P. Didot, 1817).
<https://gallica.bnf.fr/ark:/12148/bpt6k5823063x?rk=42918;4#>

Galimard, A. *Examen du Salon de 1849*, (Paris: n. p., 1849).
<https://gallica.bnf.fr/ark:/12148/bpt6k62160425?rk=21459;2>

Gautier, T., *Salon de 1847*, (Paris: n. p., 1847).
<https://gallica.bnf.fr/ark:/12148/bpt6k1091512?rk=21459;2#>

Goncourt, E. de, and Goncourt, J. de, *Études D'Art*, (Paris: n. p., 1855).
<https://gallica.bnf.fr/ark:/12148/bpt6k63810525?rk=21459;2>

Grün, A., *Salon de 1852*, (Paris: n. p., 1852).
<https://gallica.bnf.fr/ark:/12148/bpt6k873212v?rk=21459;2#>

Guizot, M., *De L'État des Beaux-Art en France, et du Salon de 1810*. (Paris: Chez Maradan, 1810).
<https://gallica.bnf.fr/ark:/12148/bpt6k65524832.r=De%20l%27%C3%A9tat%20des%20beaux-arts%20en%20France%20et%20du%20Salon%20de%201810?rk=21459;2>

Houssaye, A., 'Salon de 1844', in *L'Artiste*, May 1844, pp. 4 - 6, 17 - 20 and 53 - 57.
<https://gallica.bnf.fr/ark:/12148/bpt6k215064k?rk=21459;2>

Jal, A., *L'ombre de Diderot et le Bossu du marais: Dialogue critique sur le Salon de 1819*, (Paris: n. p., 1819). <https://gallica.bnf.fr/ark:/12148/bpt6k62305252?rk=21459;2>

Jan, L., *Le Salon de 1839*, (Paris: Beauger et Comp^{ie}, 1839).
<https://archive.org/details/lesalonde183900laur/page/n7/mode/2up>

Laincel, L. de, *Promenade aux Champs-Élysées*, (Paris: E. Dentu, 1865).
<https://gallica.bnf.fr/ark:/12148/bpt6k62160032.r=louis%20de%20laincel?rk=21459;2>

Landon, C. P., *Salon de 1808*, (Paris: n. p., 1808).
<https://books.google.co.uk/books?id=7NgUAAAAQAAJ&printsec=frontcover&dq=salon+de+1808&hl=en&sa=X&ved=2ahUKEwit-aWN8crsAhWVQhUIHTXvAXAQ6AEwAHoECAIQAg#v=onepage&q=salon%20de%201808&f=false>

Laviron, G., and Galbacio, B., *Le Salon de 1833*, (Paris: Abel Ledoux, 1833).
<https://gallica.bnf.fr/ark:/12148/bpt6k6213664w?rk=21459;2>

L'Avisé, G., *Gaspard L'Avisé au Salon de 1812*, (Paris: Chez les Marchands de Nouveautés, 1812). <https://gallica.bnf.fr/ark:/12148/bpt6k6241645r?rk=21459;2#>

Lépinos, E. de B., de, *L'Art dans la Rue et l'art au Salon*, (Paris: n. p., 1859).
<https://gallica.bnf.fr/ark:/12148/bpt6k6216873g?rk=42918;4>

Le Blanc, N. V. S. S., *Le Noir et Le Blanc, ou ma Promenade au Salon de peinture*, (Paris: Chez les Marchands de Nouveautés, 1812).
<https://gallica.bnf.fr/ark:/12148/bpt6k62277388?rk=21459;2>

Lemonnier, C., *Salon de Paris, 1870*, (Paris: n. p., 1870).
<https://gallica.bnf.fr/ark:/12148/bpt6k55038w?rk=579402;0>

Lenormant, C., *Les Artistes Contemporain. Salons de 1831 et 1833*, (Paris: A. Mesnier, 1833).
<https://archive.org/details/lesartistescont01lenogooq>

Loudon, E., *Le Salon de 1852*, (Paris: n. p., 1852).
<https://gallica.bnf.fr/ark:/12148/bpt6k8732150?rk=21459;2#>

Mantz, P., *Salon de 1847*, (Paris: F. Sartorius, 1847).
<https://gallica.bnf.fr/ark:/12148/bpt6k5817515j?rk=21459;2>

Miel, E., F., A., M., *Essai sur les beaux-arts, et particulièrement sure le Salon de 1817*, (Paris: Chez Pélicier, 1817).

<https://books.google.co.uk/books?id=SkAGAAAQAAJ&printsec=frontcover&dq=salon+de+1817+miel&hl=en&sa=X&ved=2ahUKEwiB6-Cfos3sAhUzTxUIHQjNAuEQ6AEwAHoECAUQAg#v=onepage&q=salon%20de%201817%20miel&f=false>

Montaignon, A., de, 'Salon de 1875', in *Gazette des Beaux-Arts*, (1875), Periode II, pp. 62 - 95. <https://gallica.bnf.fr/ark:/12148/bpt6k203100h?rk=42918;4>

Nau-Deville, *Sur L'Exposition des Tableau au Sallon du Louvre, 1789*, (Paris: N. H. Nyon. 1789). <https://gallica.bnf.fr/ark:/12148/bpt6k873190j>

Paturot, N., *Le Salon de 1874*, (Paris: n. p., 1874). <https://archive.org/details/lesalonde00patugooq>

Planche, G., 'Salon de 1831', in *Études sure l'école Française*, (Paris: Michel Levy Frères, 1855), pp. 7 - 173. <https://gallica.bnf.fr/ark:/12148/bpt6k6464840b?rk=85837;2>

Proth, M., *Voyage au Pays des Peintres: Salon de 1875*, (Paris: H. Vatou, 1875). <https://gallica.bnf.fr/ark:/12148/bpt6k2205716?rk=42918;4>

Robert, J., *Album du Salon de 1840*, (Paris: n. p., 1840). <https://gallica.bnf.fr/ark:/12148/bpt6k96904805?rk=21459;2>

Rochenoire, J. de la, *Le Salon de 1855*, (Paris: Martinon). <https://gallica.bnf.fr/ark:/12148/bpt6k12656940?rk=21459;2>

Sazerac, H. L., *Lettres Sur le Salon de 1834*, (Paris: n. p., 1834). <https://archive.org/details/lettressurlesalo00saze>

Stendhal, H.B., *Mélange d'art et de littérature*, (Paris: Michel-Lévy frères., 1867). <https://gallica.bnf.fr/ark:/12148/bpt6k3411112p/f361.item#>

Stevens, M. *Impressions d'une femme au salon de 1859*, (Paris: A. Boudilliat et C^{ie}, 1859). <https://gallica.bnf.fr/ark:/12148/bpt6k6215992v?rk=21459;2>

Swarte, V., de, *Lettres sur le Salon de 1876*, (Paris: Saint-Omer, 1876). <https://gallica.bnf.fr/ark:/12148/bpt6k9205276?rk=128756;0>

- Ténint, W., *Album du Salon de 1841*, (Paris: Challamel, 1841).
<https://gallica.bnf.fr/ark:/12148/bpt6k6212874n?rk=21459;2>
- Thiers, M. A., *Salon de mil huit cent vingt-quatre*, (Paris: n. p., 1824).
<https://gallica.bnf.fr/ark:/12148/bpt6k6441508m?rk=64378;0>
- Thoré, T., *Salons de W. Bürger 1861 à 1868, vol. 2: 1864 - 1868*, (Paris: Renouard, 1870).
<https://bibliotheque-numerique.inha.fr/collection/item/12889-redirection>
- Vergnaud, A. D., *Examen du Salon de 1834*, (Paris: Delaunay, 1834).
<https://gallica.bnf.fr/ark:/12148/bpt6k58187411?rk=21459;2>
- Veron, T., *Artistes de mon Temps*, (Paris: n. p., 1875). <https://bibliotheque-numerique.inha.fr/collection/item/13354-redirection>
- Vignon, C., *Salon de 1852*, (Paris: n. p., 1852).
<https://gallica.bnf.fr/ark:/12148/bpt6k6471130w?rk=21459;2#>
- Zola, E. *Mon Salon*, (Paris: Libraire Centrale, 1866).
<https://gallica.bnf.fr/ark:/12148/bpt6k64709696?rk=21459;2>