**Print Politics in the Digital Archive, 1789-99**

**Chapter 1: Introduction**

**1. Introduction**

This dissertation in concerned with the mechanism by which texts do or don’t ‘enter’ into ‘canons’ of writings, particularly in the heightened political environment of England in the 1790s. In service of this goal, this dissertation seeks to determine, in as minute detail as possible, what the print landscape in England 1789-99 was actually like. It also seeks to explain, in equally minute detail, why this task of historical recovery is, in fact, impossible. In the process, I show how our persistently perplexing questions about canons can be asked, and answered, in new ways with digital tools. Computational methods, in their close association with ‘big data’, are often advertised as the solution to the problems of scale which are presumed to motivate the formation of selective canons. As my own computational experimentation will reveal, however, computational study more commonly undertakes the same judgment and valuation of texts in ‘canon formation’ and displaces this process onto ‘corpus building.’ It is difficult, when faced with flawed and complex competing narratives, to resist the impulse to simply critique and reject all options. If we may learn any lessons from the 1790s, however, surely one must be the danger of universal paranoia.These, then, are the three scholarly conversations to which this dissertation contributes: the digital humanities, as an increasingly self-reflective set of practices; eighteenth century studies, and the challenges presented by the 1790s; and the frameworks of reparative reading within queer theory which seem to offer valuable resources for both. The remainder of this chapter will describe in more detail the relevant scholarship in all three fields, then discuss the overlaps between them which enable my work. Finally, this chapter will conclude with a description of the two computational experiments which drive the dissertation as a whole, situating them within the three fields, and providing a sketch of their development across the dissertation.

***1.1. The Digital Humanities***

The field of “digital humanities,” now several scholarly generations into its development, has entered the phase in which so many scholars are carrying out “DH” research that no universally-acceptable definition of “DH” can exist. Many definitions are fundamentally methodological. These definitions describe what might be called a “computational humanities.” Other definitions take up the new importance of digitality in daily life. This kind of work is increasingly described as “digital cultural studies.” In positioning this work as a “DH” dissertation, I might at first glance seem to belong squarely within the “computational humanities.” To be sure, much of this work is intended to apply existing computational methodologies to conduct otherwise old-fashioned literary research. Several of my contributions, too, are purely methodological in nature, in the development of new code and new guidelines for computational research using particular digital resources. However, as a fuller discussion of Johanna Drucker’s distinction between “data” and “capta” will elucidate later in this chapter, no computational research exists separately from a specific, local, culturally-defined context. Accordingly, my computational research into eighteenth century literature and my development of new computational methodologies are both inevitably tied to “digital culture.” My results often have more to say about contemporary digital archives, and the history of literary record-keeping, than they do about the eighteenth century itself. I therefore rely on an extensive body of work in cultural digital studies to examine not only the eighteenth century, but eighteenth century studies.

When telling a history of “DH,” one has many histories to choose from, all freighted with different implications for contemporary practice. Ted Underwood’s “Genealogy of Distant Reading” presents a history of “distant reading” which is not for the most part centrally concerned with computers, and therefore fundamentally distinct from concepts of “digital humanities” (Earhart). [FEMINIST HISTORY OF DH.] As these competing historiographies show, much more is at stake than mere description, when certain methodologies are named as “digital humanities.” [CONCLUSION TO THIS PARAGRAPH.]

This dissertation will, nonetheless, make use of algorithms, “data”, and data visualizations to carry out its inquiry. At every possible point, however, the underlying methodology will be made visible, and its assumptions scrutinized. Much of the code underlying this project I have written myself. Some has been written at my request. In every case where the code is available to me, the program itself appears in Appendix A (“Codebase”), accompanied by a plain-language explanation of how it operates. Where I have used closed-source software, Appendix A contains an explanation of my best guess at its underlying process. My exact use of these tools — sufficient for another to replicate my work — is provided in Appendix B (“Methodology”). These details are explicated in full in the appendices in order not to over-burden the body of the dissertation, but they are by no means *confined* to the appendices. Computation is not a “black box” to be consulted for simple answers, but an inextricable from my reasoning and argument.

My attention to the *sources* of digital knowledge creation comes from Johanna Drucker, and her distinction between “data” and “capta.” Drucker, in “Humanities Approaches to Graphical Display,” specifically addresses the digital humanities practice of creating, and then close-reading, data visualizations. She argues that the tools for visual representation which may be effective in the sciences cannot be simply and uncritically transposed to humanistic subject matter. When an experiment is presented as a ‘data visualization,’ she says, “the rendering of statistical information into graphical form gives it a simplicity and legibility that hides every aspect of the original interpretative framework” {Drucker:2011hu p.8}. In fields where the readers of such charts are also frequent creators of charts, and where norms exist to explicitly describe one’s interpretive frameworks in a methodology section, the simplicity and legibility of an individual chart may be a benefit which does not impede complex scrutiny of the information it presents.[[1]](#footnote-1) In a field like literature, however, the “graphical force” of something like a network graph or even a simple pie chart “conceals what the statistician knows very well — that no ‘data’ preexist their parameterization” {Drucker:2011hu p.8}. Drucker problematizes the term “data,” the etymology of which presents it as a “given” which is stable and independent of observation. She proposes that humanities visualizations embrace, instead, the framework of “capta,” that which is “‘taken’ actively” {Drucker:2011hu p.3}, “fundamentally codependent, constituted relationally, between observer and observed phenomena” {Drucker:2011hu p.50}. Drucker’s assessment shapes my own prioritization of qualitative and reflective computational research. The term “capta” itself has not seen uptake in subsequent digital humanities scholarship, even in cases where scholars explicitly take Drucker’s warnings to heart. Accordingly, for clarity, this dissertation will continue to use the more usual term “data” to refer to the information gathered for analysis here. However, as I integrate and compare a wide variety of data from many disparate sources, a preliminary task of my analysis is always to determine, as precisely as possible, how the information was captured and quantified.

Additionally, all of the figures presented in this dissertation are of my own design. My design praxis is informed by the work of Edward Tufte and Alberto Cairo, both of whom provide practical design advice in service of demystifying the visual rhetoric by which graphs present their arguments.[[2]](#footnote-2) Neither Tufte nor Cairo is a scholar of media studies; rather, they are professional practitioners of ‘data visualization’ who reflect critically on the assumptions of their work. Tufte’s work primarily strives to correct badly-designed data visualizations, and the dangerous decisions that bad design can lead people to. His most famous example is an analysis of the scientists’ report at NASA which led to the ill-fated launch of the Challenger space shuttle in 1986: as his extensive visual analysis argues, the scientists (untrained in graphic design) unintentionally obfuscated crucial information about the day’s launch conditions (Tufte). Tufte’s six principles of design[[3]](#footnote-3) primarily seek to guide undertrained designers away from misleading themselves. Cairo, following on Tufte’s work from the perspective of an active journalist, more often turns his attention to successful designs which mislead their audiences intentionally. His forthcoming book, *How Charts Lie*, addresses the readers of infographics with insights into visual literacy (Cairo). His preceding book, *The Truthful Art*, addresses the creators of good-faith infographics with insights into visual manipulation (Cairo). Cairo draws a distinction between “data visualization” and “infographics”: “an infographic tells the stories that its designer wants to explain, but a data visualization lets people build their own insights based on the evidence provided,” summarized more succinctly as “infographics to explain, data visualizations to explore” (Cairo). Using this terminology, my argument will proceed with infographics in the body of the dissertation as curated figures to support my argument, with fuller data visualizations available in Appendix C (“Data”) to allow further exploration. Following in both Tufte and Cairo’s footsteps, I conceive of the figures throughout this dissertation as rhetorical devices. In service of arguing honestly, therefore, my designs — in the body of the dissertation and in Appendix C — are accompanied by footnoted explanations of my design rationale.

In light of the challenges of “empiricism,” my goal in this dissertation is to produce a series of *fruitful models*. [Willard McCarty’s descriptive work on the concept of modelling VERBS modelling’s processes of abstraction.] [QUOTES FROM McCARTY] Another way of expressing this might be to borrow from the vocabulary of social science research: I would like to make discoveries about a “population” which is inaccessible to me, namely, “everything printed in England 1789-99.” Since I cannot access this population in full, I have taken a series of “samples,” which I hope to be representative; each sample can be examined as evidence of the makeup of the source population, but must also be examined in and of itself *as* a sample. [therefore, not attempting ‘accuracy’] [Rather than making claims about the 1790s each claim is filtered through / constrained by the specific model of the 1790s used to make it] Rather than producing a clear and pristine truth, my models are, in fact, quite likely to contradict each other. In these moments of contradiction, [we have the best opportunity to learn, because we can see why they contradict, and therefore why something may or may not be true.] [These opportunities are “fruitful”] [Expand on “fruitfulness”]

***1.2.*** ***The 1790s***

All of these data samples, algorithms, models, and self-reflective methodologies do have an aim beyond themselves: to recapture, in as minute detail as possible, the print landscape in England between January 1 1789 and December 31 1799. This eleven-year “decade” was a turbulent one across the Channel, encompassing the whole of the French Revolution, from the Estates General in 1789 to Napoleon’s coup in 1799.[[4]](#footnote-4) In England, these events caused strong and variously nationalist reactions in a country which had so recently lost its colonies in America and feared that a French invasion could come at any moment. [CONCLUSION: an environment which felt like it was on the edge of history]

The [world of letters? Public sphere? What do I call it?] in England responded to these events by [writing WAY TOO MUCH]. This is the decade of *Common Sense*, it is the decade of *Lyrical Ballads*; it is the decade of Hannah More, it is the decade of Ann Radcliffe; it was the age of wisdom, it was the age of foolishness; it was the epoch of belief, it was the epoch of incredulity. Charles Dickens’ now-famous superlative degree of comparison captures the tension between empiricist or otherwise political thinkers and Romantic or Gothic artists as contemporary literary scholars grapple with these modes. Each literary ‘mode’ was imbued with the potential for national importance.

I am by no means the first to examine the writing of the 1790s in light of authors’ political strategies in the print marketplace, though preceding work is often organized along genre lines. For example, George Taylor, in *The French Revolution and the London Stage*, sets forth some key contentions about literature’s cultural capital. He argues that although eighteenth-century critics sought to make distinctions in the social value of different plays, or different parts of the theatrical program at a single show, audiences did not make these distinctions. “Although audiences were divided according to class between pit, box and gallery,” he argues, the division was purely physical: “all came to see the same shows, for within one programme there was a variety that not only satisfied the different groups but satisfied them all […] all received plaudits from all corners of the house” (Taylor 3). Taylor proceeds to define the central concerns and strategies of the plays staged during this period, arguing that “the determining feature of the period for the theatre was the sense of loss, a feeling that things had changed in ways that people did not expect or want,” a feeling that was reflected in “images of incarceration, dislocation and supernatural intervention” (Taylor 221). This analysis of the theatre is structured by his originating observation that “Disagreement as to what is trash and what is treasure suggests cultural crisis, when values are put under question by social stress or political conflict” (Taylor 3). The communal space of the theatre, which some attempt to render non-communal and to dissect class lines, is for Taylor fertile ground. For my inquiry, it is the “[d]isagreement as to what is trash and what it treasure” itself which animates me: the terms “trash” and “treasure” are hardly exaggerations of the extremity of the conflicting assessments which could be passed regarding the same [works](scrivcmt://26087BCC-148F-4626-B984-5CA87FD425FF).

The desire to distinguish between “trash” and “treasure” also appears in studies of poetry, which in this period are nearly synonymous with studies of Romantic poetry. Particularly where poetry is examined in light of conflicts with France, analysis of poets’ work often doubles as an analysis of their success or failure in the pursuit of cultural capital. For example, Simon Bainbridge, in *British Poetry and the Revolutionary and Napoleonic Wars*, sees the decade and its poetry through the lens of war. His own interest is in what he terms “the attempts made by several writers to fill the role of national bard prior to Scott” (Bainbridge 3). This framework leads Bainbridge to examine Charlotte Smith alongside Samuel Taylor Coleridge, William Wordsworth, Robert Southey, and eventually Walter Scott. In his conception of ‘a national bard’ he imagines a particular marriage of poetry’s special status as imaginative literature to a specific poet’s cultivation of a nationally important public persona. Both poetry and the poet are thus pursuing a particular kind of cultural capital that allows them to rise above their own popularity. Bainbridge frequently discusses strategies of print publication as part of his analysis: a particularly insightful footnote discusses, for example, a poem which circulated both as an ode in a magazine and as lyrics in a broadsheet. Bainbridge observes that “texts could shift in status and meaning as they were adapted for different forms of production and different audiences, in this case existing in both the relatively elevated form of the ode, to be read by an individual magazine reader, and as a new set of lyrics to a traditional tune, achieving its fullest realization in a communal performance at a political meeting in an alehouse or at a demonstration” (Bainbridge 10). However, his explanations of individual print strategies are often hindered by his approach to literature’s cultural capital. At one point, for example, he argues that “the fact that the Gentleman’s Magazine was forced to reproduce at length a poem written two decades previously again emphasizes what was seen to be the failure of anyone convincingly to fill the role of the Bard until the emergence of Walter Scott” (Bainbridge 53). I contend that the republication of substantially older works of literature is not a sign that England was destitute of true poetry before Walter Scott, but rather an attempt to solve the challenges to national identity posed by the 1790s’ disruption of history.

Bainbridge is not alone is taking Walter Scott as a kind of pinnacle, and looking backwards in search of his antecedents. Richard Cronin, in *The Politics of Romantic Poetry*, concerns himself with “poetry that speaks to a divided society in an attempt to constitute its readers as citizens of […] ‘the pure commonwealth’” (Cronin 13). This definition of “political poetry” leads him to Erasmus Darwin as the poet of the the Revolution and the Revolutionary war, William Blake as the poet of the Napoleonic wars, and Richard Southey and Walter Savage Landor as the poets of the post-war period — ending, irresistibly, as so many do, with Walter Scott. Despite the fact that Cronin’s chief interest is in each poet’s use of language to construct an ideal audience, print culture looms. Of Erasmus Darwin, Cronin emphasizes Darwin’s use of “[t]he most powerful of the languages available in the 1790s that could plausibly claim universal status,” namely, “the language of science” (Cronin 16), but it is the “beautifully and expensively produced edition of The Botanic Garden” (Cronin 29) which asserts his importance. Blake’s eccentric, idiosyncratic language is mirrored in his eccentric, idiosyncratic printing practices (Cronin 60). Southey and Landor’s universal, international language seeks to “escape from history” (Cronin 81), which entails disavowing not only the popular prose fictions which sought to represent history but also the very circulating libraries in which they were distributed (Cronin 74). In other words, Cronin, even more so than Bainbridge, seems to use ‘poetic language’ to explain phenomena which have important print culture dimensions. Nonetheless, his discussion of Scott accurately describes the particular nature of Scott’s success: “Scott’s *Minstrelsy* transforms ballads freely passed from speaker to speaker around cottage fires on the Scottish Borders into luxury items, items only available to book-buyers of some means” (Cronin 94). The Romantic poets, in both Bainbridge and Cronin’s accounts, support their claims to insight and importance through the inaccessibility of their work in print.

Poetry thus stands in a stark contrast to the volumes of writing which John Mee discusses in *Print, Publicity, and Popular Radicalism in the 1790s.* Although his title leaves open the possibility that he could consider any kind of writing in print, Mee is concerned solely with the activities of the London Corresponding Society, placed in the context of Romanticism. He examines “public lectures, toasting, tavern debates, and song, but also more mundane and less colourful associational practices, such as day-to-day editorial discussion about what to publish under the LCS’s name” (Mee 28). These writers held a fundamentally different relationship to the print marketplace than the Romantic poets, or even than theatre critics: rather than concerning themselves with distinctions between ‘good’ and ‘bad’ forms of writing or publication, Mee argues, “At certain points the societies seem to operate under the spell of ‘print magic’, that is, a faith that print could liberate mankind simply by bringing ideas into printed circulation” (Mee 24). This partly shifts to Mee himself the task of assessing whether the texts he studies are ‘literature.’ A partial claim to the status of ‘literature’ is suggested by the participation of his book within the series ‘Cambridge Studies in Romanticism,’ which, he acknowledges “implies an understanding of popular radicalism as a kind of ‘literary’ culture. At least, it argues for the centrality of the writing, production, and circulation of printed texts that took up so much of the time of the radical societies” (Mee 18). Nonetheless, the book distances itself from “Romanticism and its major poets, novelists, and playwrights” (Mee 18). Mee thus replicates the distance that the period’s “poets, novelists, and playwrights” sought to put between themselves as the “the role of print personality[[5]](#footnote-5) as a form of mediation” (Mee 29) that interests Mee in the LCS.

John Guillory contends that the “canon debate” taking place during his moment of writing in 1993 “signifies nothing less than a crisis in the form of cultural capital we call ‘literature’” (Guillory p.viii); as both he and I argue, the roots of this crisis lie two hundred years earlier, in the eighteenth century formation of a vernacular (rather than classical) literary canon, and in the 1790s “first crisis” (Guillory p.xi) of that canon. Guillory turns particularly to Wordsworth and Coleridge, and the strategies by which they confer a high status on their own work by undermining the cultural capital of other forms. Wordsworth, Coleridge, and the other ‘big six’ Romantics are surely the winners of the eighteenth century historical battle for cultural capital; just as surely, the loser has been the Gothic. E.J. Clery’s *The Rise of Supernatural Fiction, 1762-1800* provides an account of what was at stake for the representations of supernatural events in supernatural stories in fiction, drama, and popular news (Clery). The “rise” she describes is not an increase in volume and prominence of supernatural stories, since her starting point in 1762 (the Cock Lane ghost) is a major national phenomenon with many imitators. Rather, supernatural fiction ‘rises’ when it acquires cultural legitimacy. Michael Gamer has more recently expanded on how this ‘rise’ fuelled Romanticism’s own rise (Gamer). In *Romanticism and the Gothic: Genre, Reception, and Canon Formation*, he details the interconnectednesss of what are now seen as the separate categories of ‘high’ Romantic literature and ‘low,’ popular Gothic writing. By using Gothic materials in self-avowedly non-Gothic ways, Gamer argues, Romantic writers could appeal to popular taste without risking the loss of cultural capital which attended the Gothic’s “popularity”. Gamer, like Guillory, primarily examines Wordsworth and the ‘winners’ of the struggle for cultural capital: I, like Clery, am more interested in the ‘losers.’ Accordingly, I attend to much that is *not* literature, in order to better understand why it is not.

***1.3. Reparative Reading***

My primary theoretical framework is that of reparative reading. Eve Sedgwick persuasively describes, in “Paranoid Reading and Reparative Reading,” the dominance of paranoia in literary criticism {Sedgwick:2015up}, and attempts to sketch an alternative in what she terms reparative reading. My touchstones are two descriptions from Sedgwick’s original chapter, “Paranoid Reading and Reparative Reading”:

The desire of a reparative impulse... is additive and accretive. Its fear, a realistic one, is that the culture surrounding it is inadequate or inimical to its nurture; it wants to assemble and confer plenitude to an object that will then have resources to offer to an inchoate self. (149)

What we can best learn from such practices are, perhaps, the many ways selves and communities succeed in extracting sustenance from the objects of a culture - even of a culture whose avowed desire has often been not to sustain them. (150-151).

[RESPOND TO THESE QUOTES.] I understand reparative reading to mean [DEFINITION OF REPARATIVE READING]. Others have worked to carry out scholarship in this mode, including [OTHER REPARATIVE READERS]. [Felski’s neophenomenology in Everyday Aesthetics] Rita Felski, in her article “After Suspicion” and then further in her recent monograph *The Limits of Critique*, continues in Sedgwick’s line. Felski asks: “Can we be postcritical — as distinct from uncritical?” (Felski 151). Reparative readers are often critiqued for painting a much clearer picture of the paranoid mode than the reparative mode, leaving it unclear how one is meant to escape paranoia’s double-bind. [WHY I VALUE REPARATIVE READING (INDEPENDENT OF DH & 1790s)].

Implicit in my prioritization of reparative reading is an affinity for queer theory more broadly. I take it as read, for example, that history consists of major discontinuities in cultural ontology and identity, as Foucault describes. I follow Judith Butler in understanding these historically-framed identities as being “performative,” actively re-created in conversation between self and others. (In my understanding of performative identity creation I also borrow from J.L. Austin, who is not typically understood as a ‘queer theorist,’ but who contributes valuable vocabulary for how words do things.)

My search for other ways of relating to literature has also led me to emerging work under the umbrella of “surface reading”. “Surface reading” positions itself as an alternative to “symptomatic reading”; [RATHER THAN X, IT Y.] The analogues to reparative and paranoid reading are obvious, but also, I argue, overly simplistic: all paranoid reading is symptomatic, but not all symptomatic reading is paranoid. Reparative reading, as described by Sedgwick and potentially modelled by Felski, is still interested in ‘deep’ meanings of texts, in which striking textual features can be interpreted to locate additional meanings. [A LOT OF QUEER THEORY IS ALSO SYMPTOMATIC, EVEN WHEN IT’S NOT PARANOID (AND TBH IT’S OFTEN PARANOID).] In contrast, “surface reading” [DOES SOMETHING ELSE.] [GIVE EXAMPLES OF SURFACE READINGS.] [GROUP/CATEGORIZE SURFACE READINGS.] [CONCLUSION]

Surface reading and reparative reading are not synonymous [modes](scrivcmt://4D050EC1-F5F0-48B4-B6EC-E6C31DBCC5BC). Felski, at the forefront of reparative reading, explicitly distances herself from surface reading: “I have no quarrel with interpretation, even though I favor description; nor am I drawn to a language of textual surfaces over depths” (CITE). The tensions between these two modes of reading are explored in more detail in A.C. Facundo’s *Oscillations of Literary Theory: The Paranoid Imperative and Queer Reparative*.

**2. Synthesis**

Drawing together three scholarly conversations with three distinct histories, it is difficult to attentively delineate the links and the tensions between each thread. Especially since many of these fields have, in fact, overlapped substantially, there is a danger of simplistically conflating distinct lines of thoughts. At the same time, over-emphasis on distinctions can overstate the separations between fields which do in fact (as the existence of this dissertation contents) share important affinities. To navigate these conceptual overlaps carefully, let us make our way through each area of an imagined Venn diagram of the three. We can discuss each pair in turn — DH and eighteenth century studies; eighteenth century studies and queer theory; queer theory and DH — to make our way toward the centre where all three overlap.

***2.1.*** ***The Digital Humanities & the 1790s***

Eighteenth century materials of various kinds have been collected in many digital archives, of very different scope. These are discussed in more detail in Chapter 2. On the tip of anyone’s tongue, of course, is Gale’s Eighteenth Century Collections Online (ECCO), containing over 180,000 titles 1701-1800, of which 42,000 were printed in England between 1789 and 1799. ECCO is itself (mostly) a subset of the broader English Short Title Catalogue (ESTC), which contains more 460,000 texts 1473-1800, of which 51,965 were printed in England between 1789 and 1799 (indicating that nearly 10,000 in the decade titles appear in the ESTC but not ECCO). The ESTC does not provide access to texts themselves: instead, it is an authoritative bibliographic catalogue, available as a searchable database. It is ECCO which provides texts themselves: ECCO’s 180,000 titles works are available as photographed facsimiles of the full text of each title. The facsimiles can be searched within ECCO’s online interface; these searches examine a plaintext version of the facsimile pages that was generated by Optical Character Recognition (OCR), but this OCR text is not made directly available. As a result, the facsimiles may be read individually by scholars, but cannot form the basis for computational corpus analysis. A subset of ECCO’s texts have been hand-prepared, as part of the Text Creation Partnership (TCP), to be easier to use in computational research. The resulting corpus of ECCO-TCP texts contains 2,231 titles, of which 466 were printed in England between 1789 and 1799. These titles are available as carefully-edited texts encoded according to the Text Encoding Initiative (TEI) standard, which not only provides an accurate version of the text’s words, but encodes substantial details regarding its context on the page. Most large-scale distant reading of eighteenth century literature relies in the ECCO-TCP corpus as its ‘model’ or ‘sample’ to represent the period. Accordingly, one of the tasks of this dissertation is to precisely examine the makeup of thus corpus, and how it differs both from other corpora and from print culture in the period itself. These three digital collections — ECCO, ESTC, and ECCO-TCP — are the primary digital resources for the period, which form the basis of most digital research. However, they represent only one approach toward the collection and presentation of digital texts, to which there are two broad kinds of alternatives. These large but meticulous collections occupy a middle space between, on the one hand, highly selective thematic collections, of which there are many, and the giants of indiscriminate textual accumulation, of which there are few.

Smaller collections allow for more scholarly curation, but have corresponding limitations. Whereas the ‘main players’ can be easily enumerated for the mega-archives, these specialized collections are numerous. Some will focus on particular kinds of texts, such as the Early Novels Database (2,041 novels 1700-1799) or Broadside Ballads Online (more than 30,000 broadside ballads). Others exhaustively index particular publications, such as The Hampshire Chronicle, the Index to the Lady’s Magazine (1770 to 1818), or the Novels Reviewed Database (1,836 reviews from The Critical Review and The Monthly Review, 1790-1820). Feminist scholarship in particular has seen the creation of resources like the Orlando Project, the Chawton House library Novels Online, Northeastern University’s Women Writers Online and UC Davis’s British Women Romantic Poets. The virtue of these collections is that they achieve even greater accuracy and comprehensiveness within their defined scope. The Shelley-Godwin Archive, for example, can reasonably aspire to digitize *every* known manuscript of Percy Bysshe Shelley, Mary Wollstonecraft Shelley, William Godwin, and Mary Wollstonecraft, and to provide these manuscripts in hand-encoded plaintext transcripts. However, as is inevitable, these specialized archives have the vices of their virtues: their specialized focus allows them to adapt precisely to their materials, and their idiosyncratic data structures can rarely be combined with other resources. The William Blake Archive, for example, benefits enormously from designing its archive around the unique images of each page of each copy of each of Blake’s works. But because this approach is so well-suited to Blake, it cannot be applied beyond Blake. Even if the archive’s resources were available for download, they could not be directly compared to materials from another source which does not record its information at such a minute level of detail.

Also in the category of ‘smaller and specialized’ archives is Project Gutenberg. Project Gutenberg makes no claims to scholarly reliability but nonetheless underlies a not-significant amount of scholarly work[[6]](#footnote-6) — its cultural capital as a resource lags far behind its use and utility. Project Gutenberg is easily conceived of as a haphazard, ‘unscholarly’ source for materials, but unlike Google Books, Project Gutenberg actually does have selection criteria. Project Gutenberg is only interested in public domain works which contemporary audiences might be interested in reading for pleasure. It narrows the field substantially to exclude works which have either ceased to be broadly interesting (as in the case of most forgotten fiction), or which was never particularly interesting (as in the case of almanacs and tax codes). Project Gutenberg includes 57,796 texts: far more than specialized scholarly archives like the Early Novels Database or the Shelley-Godwin Archive, but nonetheless an order of magnitude fewer than its more-voracious potential competitors. And, like smaller specialized scholarly archives, Project Gutenberg has tailored its holdings to make it easy for readers to read, and quite difficult for its collection to be applied to any other use. By tailoring the structure of the archive itself to its specific materials, these collections are able to thoughtfully achieve their aims — but they also make it correspondingly difficult for users to achieve their own, different aims.

The best known mega-archive is, of course, Google Books. In a scholarly context, one hesitates even to designate this as an “archive,” particularly in the same breath as resources like ECCO: books of all kinds are scanned indiscriminately with only the bare minimum of roughly-accurate metadata collected about them. Books are sourced from libraries whose collections are being scanned in bulk, or are submitted directly by publishers or authors who are attracted by Google’s call to “Promote your books on Google—for free” (CITE[[7]](#footnote-7)). These rapidly-scanned books are prone to unpredictable errors, including inaccurate dates, misspellings, duplicate copies, and inaccurate subject classifications (Harper 2016; Jacsó 2008; Weiss 2016) (CITE Mike Sutton and Mark D. Griffiths) — infamously, many books have “1899” assigned as their publication date because this date was used as a placeholder for “no date”.[[8]](#footnote-8) Many photographed pages still include the fingers of the employee holding open the book. Nonetheless, Google Books is frequently used to study the prevalence of various “n-grams” (words or short phrases) over time, thanks to Google’s built-in tool. The tool is able to search books which are, for copyright restrictions, not available directly to readers, making it highly tempting for question about contemporary language use. Similar in scope to Google Books but with more limited research tools is The Internet Archive. The Internet Archive, unlike Google Books, is a non-profit, but like Google Books, it carries out mass scanning of books. The Internet Archive declares that it contains 20 million books and texts, and scans 1,000 books per day in 28 locations around the world (CITE).[[9]](#footnote-9) Equivalent numbers are not published for Google Books, which has received less attention from Google as it encountered more legal barriers; in 2013 books were scanned at the rate of 1,000 pages per hour, and in 2015 more than 25 million books had been scanned. Google Books, Project Gutenberg, and the Internet Archive are not “canonical” sources of texts, but they are nonetheless part of the ecosystem of digital eighteenth century studies.

Moreover, as in eighteenth century texts themselves, the boundary between “canonical” and “noncanonical” digital archives is a permeable one: Google Books and the Internet Archive are also integrated into HathiTrust, an increasingly popular resource for scholars. [EXPLANATION OF HATHITRUST]. The scholarly benefit of HathiTrust is most evident in the study of contemporary literature. On a technical level, whereas OCR typically produces gibberish when faced with eighteenth century typefaces captured in photographs of microfiche, OCR is extremely effective at reading contemporary typefaces and direct high-resolution digital scans of books.

As I scrutinize the underlying assumptions that are structurally reinforced by this work, I do so not in order to dismiss their value. [Literary study already involves modelling / sampling]. Thus, I see DH and bibliography as able to enhance each other reciprocally.

***2.2.*** ***The 1790s & Queer Theory***

Queer readings of the 1790s — which is to say, of the Gothic — provide an invigorating way to value some of the most vibrant writing of the decade precisely for the features which rendered it less than “respectable.” [HISTORY OF QUEER GOTHIC STUDIES - Palmer I think, leading up to Franz Potter… this history is prob multiple paras] [A lot of really good Gothic work is connected thematically rather than strictly chronologically, examining the “transtemporal affinities and connections” (CITE) that Felski argues are often shut out by contemporary historicist critique. And indeed, this mode of study is falling out of favour regarding the Female Gothic.]

[A lot of queer Gothic work has been paranoid] [Something about Susan Sontag & camp being how I reparatively read the Gothic]

Although the queerest work in this period has centred on the Gothic, I intentionally do not limit my focus to the Gothic. My broader scope is, in part, based in an argument about the Gothic’s prevalence: following scholars like [EXAMPLE], I see the Gothic not as a clearly-defined literary genre, but rather as an affective ‘mode’ that can (and did) permeate writing far beyond the realm of supernatural fiction. Accordingly, [although I am interested in the Gothic, I study everything.] [Also I’m not just interested in the gothic.]

***2.3.*** ***Queer Theory & the Digital Humanities***

My emphasis on transparent, critical, and reflective praxis in the capture and visual presentation of data[[10]](#footnote-10) owes much to the emerging field of critical algorithm studies. Any methodology is, to a certain extent, an “algorithm,” in the loose definition of ‘a series of pre-defined steps to be carried out’. But computational algorithms differ from “algorithms” implemented by humans. Computational algorithms have two key vulnerabilities: first, their operations are less easily scrutinized; second, their results are more easily trusted. The second vulnerability — the cultural aura of empirical trustworthiness which accrues to anything ‘computational’ — is another flavour of the same vulnerability that Drucker describes with ‘data’ generally. Because the human agents who designed and trained any given algorithm appear to be absent from its operation, the algorithm appears able to discover truth directly. This is how Daily Wire reporter Ryan Saavedra was able to tweet with disdain that “Rep. Alexandria Ocasio-Cortez (D-NY) claims that algorithms, which are driven by math, are racist” (Saavedra): anything “driven by math,” he assumes, must be incapable of human fallibilities like racism. But as Safiya Noble shows extensively in *Algorithms of Oppression*, algorithms by default reproduce, and can easily exaggerate, the assumptions and biases of the culture in which they are made (Noble). In other words, in a racist world, algorithms *are* racist — and sexist, and duplicative of all other systemic institutional inequities.

Wendy Chun suggests that the problem with much algorithmic inquiry lies in its most core assumption: the insistence that the future must look like the past (Chun). This assumption is essential to any system which exclusively bases its predictions on past data. It is particularly intertwined with what Chun calls “homophily,” “the axiom that similarity breeds connection—which grounds contemporary network science” (Chun 60). Homophily, as it is currently applied to network science, Chun argues, both “assumes and creates segregation” (Chun 76) in a fashion that allows it to operate “like money laundering for bias” (Chun 62-63). Importantly, however, the analogy to money laundering is imperfect: it cannot be resolved with “better, cleaner data” (Chun 64). Instead, Chun insists, “new theories of connection—which do not presume a dangerously banal and reciprocal notion of friendship—are needed” (Chun 89), a call to action which places its emphasis not on the information provided to an algorithm but on the algorithm’s procedures.

An additional and perhaps-unexpected resource, for scrutinizing algorithms critically, comes from the work of those who study games, especially video games. [Ian Bogost argues that procedures contain *arguments*.] This framework allows us to see that all research methodologies, including non-digital ones, contain implicit arguments about the nature of what is studied. [EXAMPLE PROCEDURE FROM MY MATERIALS - something non-computational?]

As I encounter the limitations of the various information and tools with through which I attempt to understand the 1790s, my goal is to do something other than facilely observe that they are limited. Critique, in many venues, still has valuable new information to contribute regarding the contemporary digital landscape. Drucker, Noble, and Chun’s warnings about the subjective nature of data and algorithms, for example, all [HAVE REAL AUDIENCES]. In part because their critiques have been presented so effectively, however, I do not give critique a central role here. Most of the digital artefacts that I examine have been created, not by ever-hungering capitalist corporations, but by my fellow scholars. In the spirit of laying down the next few stones in the building of a cathedral begun long before me, [I WANT TO BUILD]. In a digital humanities context, a focus on building connections can be mundanely practical: typing indexes from print works into spreadsheets, [DATA CLEANING], writing programs to process metadata: all of these maintain the functional usability of existing resources in new contexts. [But not just maintaining them, repairing them.]

By marrying “surface reading” and “distant reading” I may seem to leave “reparative reading” behind: after all, Felski’s call for postcritical analysis explicitly insists, “Sometimes serious thinking calls for a judicious decrease rather than and increase of distance” (CITE). Felski then defines a decrease of distance as “a willingness to acknowledge and more fully engage our attachments”: it is in this readiness to acknowledge and embrace the [DETAILS OF SOMETHING-SOMETHING]

[“Queer DH” is now a vibrant field in its own right.] [WRITE UP HISTORY / BROAD CATEGORIES OF QUEER DH RESEARCH] I draw on this work in my pursuit of less “extractive” and more anti-oppressive digital humanities methodologies.

Thus far, the importance of queer theory to DH is evident. But what about the other direction? Just as, I argue, both DH and bibliography have important resources to offer to each other, so to do I argue that DH has something to offer to the elements of queer theory that are concerned with reparative reading. Specifically, a well-formulated DH praxis can introduce a new relationship to the experience of *surprise* which is so challenging in paranoid reading. The most common dismissals of DH research are attempts to say, “we already knew that” — essentially, [PARANOID AVOIDANCE OF UNPLEASANT SURPRISE FROM DH THREAT]. This reaction invites two responses. First, [if the experiment told us that everything we knew was COMPLETELY WRONG, this would not be very good because probably at least some of what we know is right]. [Experiments fall into two categories: those which tell us about our methods, and those which tell us about our fields of study]. Second, as scholars such as Ted Underwood have noted, it is often only *retroactively* that these discoveries are declared to be “obvious” results which everyone “already knew” (CITE). If we are asked to form an explicit hypothesis in advance, we might be very surprised by the exact numbers. [CONCLUSION]

**3. Experiment Design**

Having delineated a large and complex scholarly context for this work, and made some preliminary efforts toward synthesis of these disparate backgrounds, it is time, now, to delineate the research born of this context. This dissertation proceeds in two major parts. The first examines the 1790s through the lens of ‘titles,’ and the second through ‘persons.’ In both cases my central interest is the vexed category of ‘popularity.’ With both ‘titles’ and ‘persons’, I establish the materials with which I will be working and their implicit models. Using the information they make most readily accessible, I explore which works and individuals emerge as ‘popular’ or ‘important’ in each resource. I also probe each resource for its representation of (or failure to represent) my touchstone authors, Ann Radcliffe, Charlotte Turner Smith, Hannah More, and Mary Robinson. The resulting findings are primarily used to evaluate how the underlying digital resources participate in or resist the projects of canon-building. Then, for both titles and persons, there is a ‘turn.’ For each area of inquiry I introduce a novel experiment which synthesizes my materials and repurposes them. These experiments are then used to draw conclusions about 1790s print culture.

***3.1. Titles***

The first phase of this project takes up titles and databases. Throughout this work, a ‘title’ is the broadest possible umbrella term for a text: it is synonymous with ‘a database entry which I believe to represent a printed work.’ Each individual database will shape the precise correspondences between a ‘title’ and the printed texts about which I use ‘titles’ to reason. For example, it will vary based on my source whether multiple editions of the same work are several ‘titles’ or one ‘title,’ and whether the titles exclusively represent ‘books’ or also include chapbooks, broadsheets, or even ephemera. Due to the inscrutability of digital infrastructure discussed above, it is often not possible for me to know prior to experimentation what kinds of titles I am working with. Chapter two, therefore, two takes up contemporary digital archives directly, examining corpora of eighteenth-century literature through the same critical lens by which anthologies and classroom teaching are often scrutinized.

In chapter two, I make the case that digital archives can implicitly shape scholarly research, and begin the process of revealing and interrogating their invisible assumptions. The chapter begins with a task somewhere between a literature review and a scientific meta-analysis. My first goal will be to survey as broadly as possible the accessible mass holdings of eighteenth-century texts (all those containing at least 100 works from the 1790s). I expect to find systematic exclusions where archives are investing more labour in their holdings, with narrower selections as they move from bibliographic data to facsimiles to scholarly transcripts. To contextualize these decisions about inclusion, I research the history of how each corpus was formed. As part of this demystification process, I also discuss and theorize the difficulties involved in researching these histories: drawing on, for example, my experience with HathiTrust’s codebase, I critique the assumption that digital resources make all information transparent and accessible. Returning to the actual contents of each archive, I discuss the nature of their exclusions, and consider paths to greater inclusivity. The second chapter thus establishes the corpora which will drive my argument in chapter three, and will shape the later phases of my research in chapters four and five.

In chapter three, I synthesize these disparate sources of texts and metadata, to examine popularity as it manifests in print culture. I begin by creating my own ‘superset’ database of all records. I then query this database through random sampling and topic modelling. Both random sampling and topic modelling, as methodologies, allow me to examine information outside of its pre-designed data structures. Random sampling, which Steven Zwicker (Zwicker) and Leo Lahti, Niko Ilomäki, and Mikko Tolonen (Lahti, Ilomäki, and Tolonen) have applied successfully to the English Short Title Catalogue’s holdings, makes it feasible to apply a greater level of human scrutiny to a large body of texts. Topic modelling makes it feasible regularize otherwise-disparate materials without imposing a pre-defined ontology. My use of topic modelling differs somewhat from its typical use in literary distant reading, which generally applies topic models to the full text of literary works, and examines topics themselves as proxies for subtle elements of textual content. I model only the titles of works, taking advantage of the eighteenth century’s distinctively rich title conventions.[[11]](#footnote-11) Both of these methods allow me to use my ‘superset’ database to ask questions about 1790s literature itself. Influenced by Lesser and Farmer’s articulation of “structures of popularity,” I consider popularity in terms of total number of editions, frequency of reprinting, and market share. After presenting my proposals for how to calculate each of these metrics, I ask: what was most popular during the decade, according to my corpora? How do the corpora differ in their answers, and why? Many of the most reprinted works substantially pre-date the 1790s in their composition. Accordingly, taking up David Brewer’s challenge to account for the increased “footprint” of some texts beyond the moment of their original publication, I also pay attention to works originally written before the 1790s which nonetheless can be considered important “1790s literature” due to prominent reprinting. This inquiry’s first question is one of discovery: what works resurface in the 1790s? Its next question is one of close-reading and historical context: what makes them seem newly relevant? The chapter as a whole, then, presents a sustained study of the relative popularity of the most prominent works printed during the 1790s, and seeks to answer how these prominent works might affect what we define as “popular literature.”

***3.2. Persons***

The second phase of the project takes up persons and networks. As in the first phase of the project, it is a substantial undertaking simply to recover contemporary printing practices; this time, rather than asking what was printed, I ask who it was printed by. A great deal of scholarly work already exists on printing circles, coterie publishing, and individual publishing houses. My project consults this scholarship to extract and encode connections between authors, printers, and publishers (but not patrons, readers, or other persons not immediately involved in the production of printed texts) in order to synthesize the implicit social networks underlying 1790s print production.[[12]](#footnote-12)

Chapter four introduces my mapping of the social world of print production 1789-99. My understanding and use of network analysis is contextualized in light of Wendy Chun’s work on networks in “Queerying Homophily” (Chun). I describe my methodology and its assumptions, and provide a rich description of my resulting network graph. The graph I create may show one large interconnected network, or several separate networks of varying sizes; these networks may show highly distinct clusters, or evenly interconnected webs. Drawing on mathematical graph theory, the chapter will explain the implications of whichever shape the network ultimately displays. It will also present an overview of the people I identify as the “major players” in the publishing world of the 1790s, both mathematically (looking for nodes with various kinds of centrality) and in the scholarship.

Having recaptured these complex networks in some depth, I can then examine them, in chapter five, for their relation to our current understanding of mainstream and radical--or as I am terming them “mainstream” and “non-mainstream”--printing circles. My network graphs model individual political affiliation as a complex, socially defined practice rather than a set of concrete and stable ideological stances. This chapter will look for traces of affiliation in the print practices of publishers and authors. I consider individual printers with political allegiances, as in Dissenting societies, radical publishers, and correspondence societies. This then enables me to consider authors’ strategic choices as they publish with different printers. Having identified radical elements in the publishing world, I interrogate the radicals’ claims to marginalization. I suspect that I might find that they were not as socially estranged from the mainstream as they describe themselves, and that their printed works may accordingly have been less marginal. I discuss alternative print markets and alternative circulation, in context with the print production which they are alternatives to. The circulation of works in manuscript presents me with two challenges which will be discussed here. The first challenge is methodological: the circulation of manuscripts clearly occurred, and may have constituted “publication” within social circles, but manuscripts fall outside my purview. This chapter will therefore discuss the nature and rough shape of the gap which the exclusion of manuscript works leaves in my study. The second task of this chapter is more theoretical: as queer and decolonial DH scholars note, there is an ethical choice implicated in the decision to systematically discover, collect, and expose communities which intentionally operated below the notice of state observation. Historical distance prevents me from worrying about causing direct harm through my work, but nonetheless I critically interrogate my own research practices and contextualize my choices with the horizon of expectations within the radical circles I expose. Finally, having discussed the networks of radical and mainstream publishing in the 1790s, I also compare the position of radical publishers in the 1790s with their status in the corpora discussed in chapters two and three, where they may in fact be marginalized. Together, these approaches further complicate the story of popularity which the dissertation challenges elsewhere, by suggesting ways to reassess of the popularity of radical works.

**Works Cited**

Bainbridge, Simon. *British Poetry and the Revolutionary and Napoleonic Wars*. Oxford University Press, 2003. Print.

Cairo, Alberto. “Infographics to Explain, Data Visualizations to Explore.” *The Functional Art Alberto Cairos Weblog on Visualization, Infographics, and Data Journalism*. N.p., 16 Mar. 2014. Print.

Cairo, Alberto. *The Truthful Art: Data, Charts, and Maps for Communication*. San Francisco: New Riders, 2016. Print.

Chun, Wendy Hui Kyong. “Queerying Homophily.” *Pattern Discrimination*. Lüneburg: meon press, 2018. Print.

Clery, E J. *The Rise of Supernatural Fiction*. Print.

Cronin, Richard. *The Politics of Romantic Poetry*. 2000. Print.

Drucker, Johanna. “Humanities Approaches to Graphical Display.” *Digital Humanities Quarterly* 5.1 (2011): 1–23. Print.

Earhart, Amy E. *Traces of the Old, Uses of the New: the Emergence of Digital Literary Studies*. Vol. 11. University of Michigan Press, 2017. Web.

Felski, Rita. *The Limits of Critique*. University of Chicago Press, 2015. Print.

Gamer, Michael. *Romanticism and the Gothic*. Print.

Guillory, John. *Cultural Capital: the Problem of Literary Canon Formation*. University of Chicago Press, 1993. Print.

Lahti, Leo, Niko Ilomäki, and Mikko Tolonen. “A Quantitative Study of History in the English Short-Title Catalogue (ESTC), 1470-1800.” *LIBER Quarterly* 25.2 (2015): 87–31. Web.

Mee, Jon. *Print, Publicity, and Popular Radicalism in the 1790s: the Laurel of Liberty*. 2016. Print.

Noble, Safiya. *Algorithms of Oppression*. NYU Press, 2018. Print.

Sedgwick, Eve Kosofsky. “Paranoid Reading and Reparative Reading: Or, You're So Paranoid, You Probably Think This Essay Is About You.” *Touching Feeling*. Durham & London: N.p., 2003. 123–151. Print.

Saavedra, Ryan. “Socialist Rep. Alexandria Ocasio-Cortez (D- NY) Claims That Algorithms, Which Are Driven by Math, Are Racist.” Tweet.

Taylor, George. *The French Revolution and the London Stage, 1789-1805*. Cambridge: Cambridge University Press, 2001. Web.

Tufte, E R. *The Visual Display of Quantitative Information*. 2nd ed. Graphics Press, 2001. Print.

Zwicker, Steven N. “Is There Such a Thing as Restoration Literature?.” *Huntington Library Quarterly* 69.3 (2006): 425–450. Web.

1. It may also be the case, of course, that even fields with a long history of graphical display would benefit from greater scrutiny of the evidence they use; see: the Data Dinosaur. But this is beyond the remit of what an English PhD can address. [↑](#footnote-ref-1)
2. I cite Tufte and Cairo as the thinkers whose design philosophies best accord with my own current understanding of the work and craft of persuasive data visualization, but my actual practical training as a graphic designer is indebted to Judith Galas, Sonia Davis Gutiérrez, and Tom Hapgood. [↑](#footnote-ref-2)
3. 1. show comparisons, contrasts, differences. 2. show causality, mechanism, explanation, systemic structure (intervention relies on manipulable causality -- can't do anything with the information without causality). 3. show multiple variables (3 or more) -- the world is multivariate. 4. *completely integrate* words, numbers, maps, graphics, etc. Provide information at exact point of need. 5. documentation must thoroughly describe evidence and its sources, provide complete measurement scales. 6. presentations succeed based on their content. for better presentations, get better content. [↑](#footnote-ref-3)
4. Although these events, of course, did not occur on January 1 or December 31, respectively, the entirety of 1789 and 1799 are both included in my study here, out of sheer technological necessity. [↑](#footnote-ref-4)
5. By this, Mee means “both attacks on personalities by the radical press but also the development of personae by writers and booksellers” (Mee 29). [↑](#footnote-ref-5)
6. I have heard it quipped more than once in digital humanities gatherings that you always think you’re going to get your texts from somewhere else, but Project Gutenberg is where you’ll actually get them. [↑](#footnote-ref-6)
7. https://books.google.ca/intl/en/googlebooks/partners/ [↑](#footnote-ref-7)
8. CITE http://languagelog.ldc.upenn.edu/nll/?p=1701 [↑](#footnote-ref-8)
9. https://archive.org/about/ [↑](#footnote-ref-9)
10. As a recurrent Freudian slip, I have more than once typo’d the word “data” as “danger” [↑](#footnote-ref-10)
11. For example, a typical title in the English Short Title Catalogue for the decade is “The injustice and impolicy of the slave-trade, and of the slavery of the Africans: illustrated in a sermon preached before the Connecticut Society for the Promotion of Freedom, and for the Relief of Persons Unlawfully Holden in Bondage, at their annual meeting in New-Haven, September 15, 1791. / By Jonathan Edwards, D.D. Pastor of a church in New-Haven. ; To which is added, A short sketch of the evidence for the abolition of the slave-trade, delivered before a committee of the British House of Commons.” The usefulness of this title, as an advertisement for (if not necessarily an accurate representative of!) the work’s content, I argue, presents an opportunity unique to the eighteenth century. The eighteenth century is often dismissed as unsuitable for computational distant reading because eighteenth century printed type produces very low-quality OCR transcripts — but a focus on OCR-based distant reading itself seems to me like a response to the poverty of paratextual material in later literature. [↑](#footnote-ref-11)
12. I begin my research for this chapter by encoding only a few existing studies, in order to evaluate the feasibility of my method at scale. It is possible that, rather than directly consulting the more richly historically-informed work of other scholars, I will instead fall back on inferring networks from the author and publisher metadata included with the corpora examined in chapters two and three [↑](#footnote-ref-12)