**What We Quote: Scholarly Quotation Conventions and the Atmospheres of Disciplinary History**

Sierra Eckert and Milan Terlunen

# **Abstract**

This essay studies quotations ––in particular, quotations of primary source texts in modern literary scholarship––and argues that such forms of scholarly evidence can shed light on collective formations within the history of literary study. Drawing on a body of scholarly writings published in post-1945 Anglo-American academic journals, we re-tell histories of the discipline through patterns in passages that scholars quote. In doing so, we offer a quantitative approach to disciplinary history. Taking George Eliot's novel \*Middlemarch\* (1871-2) as a case study, we employ computer-assisted methods of reading, using an algorithm to detect quotations from the novel within the body of scholarly articles held in JSTOR. What such quantitative methods reveal are the aggregate patterns in the body of quotations—what we call “textual atmospheres”––that produce and are produced by conventions of quotation. Reading patterns in quotation, we show how gradual changes and persistence in scholarly attention to specific parts of texts operates at levels beyond the individual scholar.

\*\*\*

> There is a \*\***well-known**\*\* passage in \*Middlemarch\* […] "If we had a keen vision and feeling of all ordinary human life, it would be like hearing the grass grow and the squirrel's heartbeat, and we should die of that roar which lies on the other side of silence”[@AustenWhyFeministCritics1976, 560]

> …Eliot’s \*\***famous\*\*** observation, “If we had a keen vision and feeling of all ordinary human life, it would be like hearing the grass grow and the squirrel's heartbeat, and we should die of that roar which lies on the other side of silence”[@FenvesExilingEncyclopediaIndividual1987, 431, n.10]

> “If we had a keen vision and feeling of all ordinary human life,” Eliot writes in a \*\***famous\*\*** sentence, “it would be like hearing the grass grow and the squirrel's heartbeat, and we should die of that roar which lies on the other side of silence”[@RobbinsSweatshopSublime2002, 88]

> …\*Middlemarch\*’s more \*\***famous\*\*** worry that “[i]f we had a keen vision and feeling of all ordinary human life, it would be like hearing the grass grow and the squirrel's heartbeat, and we should die of that roar which lies on the other side of silence”[@GreinerSympathyTimeAdam2009, 304]

> …\*Middlemarch\*’s \*\***famous**\*\* evocation of the social sublime: “If we had a keen vision and feeling of all ordinary human life, it would be like hearing the grass grow and the squirrel's heartbeat, and we should die of that roar which lies on the other side of silence” [@RosenthalLargeNovelLaw2010, 795]

Quotations perform many routine tasks in modern literary scholarship. They serve as a bulwark in the eyes of scholars like Cleanth Brooks against the “heretical” act of paraphrase.[@BrooksWellWroughtUrn1947] They serve to incorporate pieces of text that are not the writer’s own into their scholarship. And, even more routinely, they serve as markers of shared disciplinary knowledge. In the epigraphs above, five scholars introduce the same passage from George Eliot’s \**Middlemarch\** (1871-2), each doing so in a way that reproduces both the text from the novel and their own language marking the passage’s status as “well-known” or “famous.” While central to the everyday work of literary scholarship, quotation is mostly the subject of handbooks and style guides, not disciplinary histories. Histories of literary studies have typically been told as stories of differentiation, where the disciplinary shifts emerge through differences in the scale and choice of object.[For a more extended discussion of this history-of-the-discipline as conflict and differentiation, see @GraffProfessingLiteratureInstitutional2008; @EnglishShiftingScalesLiterature2016] By contrast, quotations, especially of the same “well-known” passage, are quiet reminders of the parts of scholarly writings that are shared or undifferentiated. As bits of shared text, quotations are sites within literary scholarship where individual differences fall away.

This essay studies quotations ––in particular, quotations of primary source texts in modern literary scholarship––and argues that such forms of scholarly evidence can shed light on collective formations within the history of literary study. Drawing on a body of scholarly writings published in post-1945 Anglo-American academic journals, we re-tell histories of the discipline through patterns in passages that scholars quote. In doing so, we offer a quantitative approach to disciplinary history. Taking George Eliot's novel \*Middlemarch\* (1871-2) as a case study, we employ computer-assisted methods of reading, using an algorithm to detect quotations from the novel within the body of scholarly articles held in JSTOR. In reading patterns in quotation, we argue that a quantitative study of scholarly quotation patterns helps reveal the role of collective knowledge as part of the history of literary scholarship.[In deploying quantitative analysis of scholarly writings to understand disciplinary history, our method is indebted to scholarship on literary canons. For foundational work see @GuilloryCulturalCapitalProblem1993, 27; @GatesLooseCanonsNotes1992; and @SmithContingenciesValueAlternative1991, 42-53. For an analysis of citations across the MLA Bibliography, booklists on social media site Goodreads, and syllabi in the Open Syllabus Project, see @BourrierSocialLivesBooks2020.]

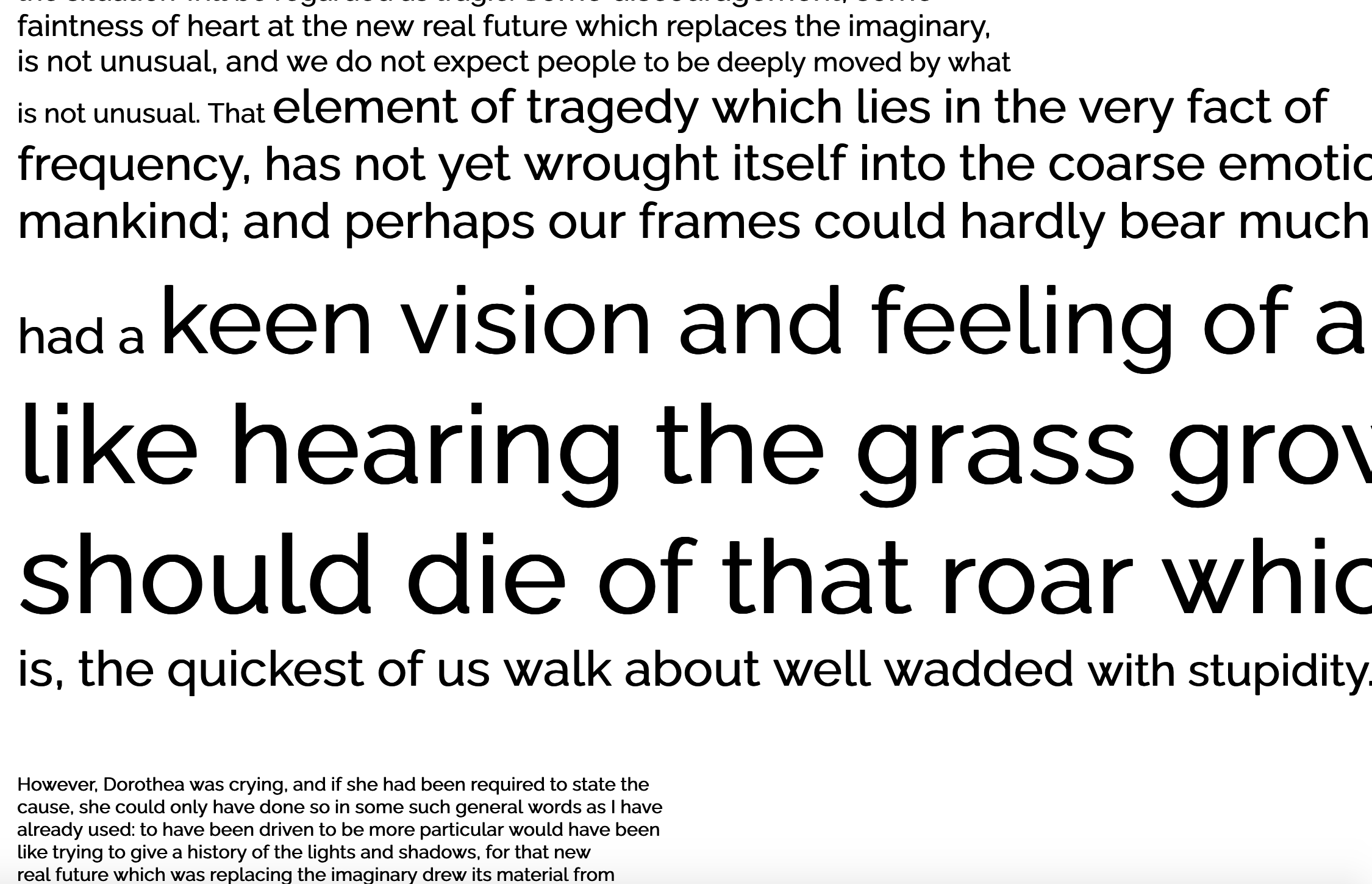
For literary scholars, quotations are fundamental units of argument––so fundamental that it can be easy to take their presence for granted. They form one part of the “citational practice” that Eric Hayot sees as a crucial scholarly activity, a practice that can “reflect in a conscious or unconscious way your belonging to a school or schools of thought.”[@HayotElementsAcademicStyle2014, 152. Hayot’s expansive definition of citational practice includes note, footnotes, quotations, and bibliographic references.] A long tradition of bibliometric scholarship, including recent feminist approaches by Sara Ahmed and Christen A. Smith (the anthropologist behind the #CiteBlackWomen campaign), have recognized citational practice––often in the form of bibliographic citations of the titles of works or names of other scholars––as concrete traces of institutional conventions, structures and inequalities within the academy.[See @AhmedLivingFeministLife2016; @SmithCiteBlackWomen.] Unlike citations, quotations––by which we mean the verbatim replication of other texts––are not as easily mapped onto institutional configurations. [There is fairly extensive scholarship on quotation as a device within \*literary\* texts, e.g. @GarberQuotationMarks2003; @CompagnonSecondeMainOu1979; @MeyerPoeticsQuotationEuropean2015; @MorsonWordsOthersQuotations2010. For a summary of critiques of the convention in the natural sciences not only not to quote but not even to cite page numbers, see @SwordStylishAcademicWriting2012, 144. A few scholars have studied the quotation practices of literary critics and reviewers in the nineteenth century, before the professionalization of academic literary studies in twentieth-century ––see @PriceAnthologyRiseNovel2000 and @DamesNotCloseReading2010]

At its most basic, quotation is an act of textual reproduction: it reproduces a piece of a text––a novel, a poem, an ephemeral genre––within a new text.Scholar’s quotations do work: aas Elaine Auyong has observed, they perform a kind of bricolage, “gathering together far-flung passages of the text and considering them side by side” in order to produce “a new configuration of verbal information.” [@AuyoungWhatWeMean2020, 99] Jonathan Kramnick describes this aspect of quotation in literary scholarship as a kind of “craftwork” as the in-line quotations often fold in and rewrite the original language of the source text.[@KramnickCriticismTruth2020, 220-221] Unlike Kramnick, our interest is not in the practice of individual scholars––how scholars quote––but in the more fundamental, yet understudied question of “what” scholars quote. Understanding how scholars quote and how those patterns persist or change over time can provide one proxy for moving from “reading” as a generalized, abstract phenomenon to the specific kinds of objects and knowledge that scholarly readings produce. When studied in aggregate, quotations of the same literary work by many different scholars produce an object that emerges only in the paper trail of scholarly writing.[In studying the artifacts produced by literary scholars, we build on studies of other institutional artifacts, like syllabi, curricula, and most recently, archival teaching materials in @BuurmaTeachingArchive2020] This object is real but diffusive: its material components are dispersed across a multitude of scholarly texts. We call this diffusive object “textual atmosphere,” by which we mean the conventions about which passages scholars frequently quote, and the material sites where they are quoted. By focusing on this aggregate, diffusive phenomenon, we offer an approach to disciplinary history grounded not in individual scholars but in collective conventions. [Our study thus shares the spirit of other quantitative approaches to disciplinary history, @GoldstoneQuietTransformationsLiterary2014, @WellmonPublicationPowerPatronage2017, @TenenVisualQuantitativeApproachesIntellectual2017].

# **Disciplinary History by the Numbers, or Thirteen Ways of Looking at Scholarly Quotations**

Studying quotation forces literary scholars to reckon with how little we know about what and how scholars quote. Take our opening example of the passage from \*Middlemarch\*, chapter 20: “If we had a keen vision and feeling of all ordinary human life, it would be like hearing the grass grow and the squirrel’s heart beat, and we should die of that roar which lies on the other side of silence.”[@EliotMiddlemarch2012, 194] What if you wanted to know how other scholars have used the passage? You might find yourself overwhelmed by a roar of textual information. The five quotations in our excerpts are just a small fraction of the works quoting this passage, which is one of the most frequently quoted passages in the novel. This exact sentence has been quoted over 50 times in scholarly articles written in the last six decades. But in order to get a sense of the critical afterlife of this passage, a scholar would need to navigate the same difficulty that Eliot’s narrator so vividly evokes: balancing between a wide, bird’s eye view and the overwhelming “roar” of all that aggregate knowledge.

To start to get a sense of how this passage has been used, we might visualize it. In Fig.1 below, we’ve resized the font size to be proportionate to how frequently each word has been quoted. Within this one paragraph, the disproportion is enormous: the clauses from “hearing the grass grow” up to “roar which lies on the other side of silence” have been quoted many more times than the clauses and sentences surrounding them. (This chart deliberately conflates many scholars together: it’s impossible to perceive individual scholars’ choices at this level of aggregation, and indeed it’s even impossible to see where each scholar’s quotation begins and ends - the changes in font size represent broad patterns, but individual scholars’ quotations may cut across these boundaries.) In short, it raises the question of how to make aggregate phenomena understandable and perceivable.



This is not just a question for \*Middlemarch\*: all literary scholars produce a selection of quotations that stand in for their literary objects of study. When writing about longer texts, especially novels (and especially a novel as long as \*Middlemarch\*), literary scholars can quote only a tiny proportion of the entire text.[This isn’t just a question of length. Short lyric poems might not seem to demand excerpting, but see @NgaiOurAestheticCategories2015, chapter 1 on how the “smallness” of short poems relates to both quotability and copyright enforcement. See also @PrinsVictorianSappho1999 and other work within the field of “historical poetics” for studies of historically-specific reproduction of poetic texts.] This creates a familiar problem of synecdoche––on what basis does a part stand in for the whole?[At the start of an analysis of Henry James’s \*The Golden Bowl\*, Martha Nussbaum teasingly states “I presuppose the quotation of the entire novel”––it’s precisely because scholars cannot presuppose this that the passages they quote from novels, or any long text, become significant, @NussbaumLoveknowledgeessays1990, 149] We take quotations to be a loose proxy for scholarly attention: although quotations aren’t the only way scholars attend to texts (think of plot summary, or generalizations about theme or character), but they are an important way that literary scholars carve up a text, using its parts in ways that, as we will show, are unevenly distributed over time.

We set out to see what we could learn about what parts of novels have been quoted at a large scale and over a long sweep of disciplinary history: the literary scholarship published in Anglo-American journals in the, a period marked by the emergence of many professional scholarly journals and bookended by shifts in the institutional infrastructure of literary scholarship. Because our questions about disciplinary history of literary study were grounded in quotation of particular works in particular institutional contexts (academic journals), we turned to JSTOR, whose holdings, while by no means a complete collection of scholarship, include the full back-catalogs of long-running journals for the discipline of literary studies (PMLA, MLQ and Modern Philology) as well as the back-catalogs of the major journals founded around mid-century as part of the institutionalization of novel studies and Victorian Studies.[^Jstor] In focusing on George Eliot’s \*Middlemarch\* (1871-2), we chose a single hyper-canonical novel that is both frequently quoted and one that, by virtue of its length, would exemplify the problems of quotation from long texts.

[^Jstor]: JSTOR is important both as a source of complete runs through 2015 of \*ELH\*, \*PMLQ\* and (representing a wider range of literary studies journals than competitors like Project Muse) and because it is one of the longerpro In addition, and putting aside questions of representativeness, focusing on one scholarly repository allows us to zero in on patterns of scholarly quotation at the level of the quotation, the individual journal, and the particular repository in which academic journals are aggregated and accessed.

Through \*JSTOR\*’s Data for Research academic portal, we were able to generate a corpus of all twentieth and twenty-first century articles including the word “Middlemarch” contained in its database.[^9] Alongside this corpus of scholarship, we downloaded the Project Gutenberg text of \*Middlemarch\*, which is a reliable plain text source.[While Project Gutenberg has relatively high transcription quality, it does not include bibliographic information about the edition consulted. For \*Middlemarch\*, a text with relatively minor editorial variants, this seemed to us an acceptable trade-off.] Next, together with our colleague Jonathan Reeve, we created a custom text-matching algorithm to compare the text of the articles in the corpus to the text of *\**Middlemarch\*.[For full text matcher code, see @ReeveJonathanReevetextmatcher2021] This allowed us to identify instances of what computer scientists call “text reuse”––that is, moments where scholars directly quote parts of the novel.[See the recent article by @PiperMeasuringUnreading2020, who use this text-matcher developed as part of a collaborative project by Jonathan Reeve, Milan Terlunen, and Sierra Eckert to analyze the citational afterlife of Goethe’s work.] We built a degree of “fuzziness” into our text matcher: first using a “stemmer” algorithm to remove all grammatical endings from words (since scholars frequently use square brackets to change these), and second building in tolerance for one character variation per word (to accommodate errors due to OCR (optical character recognition) scanning, which can mistake “1” for “l” or “rn” for “m” as well as variants between British and American spellings like “labour” vs “labor”). [We used the Lancaster stemmer, see @NltkStemLancaster. In technical terms, we allowed for character variation based on a Levenshtein distance of 1. For more technical details on our text matcher, see @middlemarchcriticalhistories.]

[^9]: In the field of corpus linguistics, the term “corpus” refers to a large and structured collection texts. Like our study, corpus linguistics identifies patterns of usage imperceptible to individual speakers and yet consequential for the language.

Our text matcher algorithm output a large amount of data on scholarly quotations from *\**Middlemarch\*. With it, we identified 482 articles written between 1960 and 2015 that quoted some part of Eliot’s novel, and collected related metadata: the length of the passage quoted, its location within the novel along with other metadata about that article: what journals these articles were published in, the years they were published, and the article’s title. We should say at the outset that we were not able detect all quotations from \*Middlemarch\* within our corpus of JSTOR scholarship. In some cases the OCR was so bad that no amount of fuzziness would have allowed for a match.The algorithm that we use works by detecting three overlapping three-word phrases that match the text of \*Middlemarch\*. It would not, for example, pick up the quotation of a short phrase such as “squirrel’s heartbeat” (unless surrounded by other words quoted from the same passage). What this means is that in our method, we’ve shied away from making claims about absence of quotation––since this may not actually reflect a total absence––and instead, we’ve focused on passages that are quoted––especially those quoted frequently.

To find frequently quoted literary passages is not to assume that these passages are quoted for the same reasons. Quotations can serve many purposes. For scholars in any discipline, a quotation may be a performative display of knowledge, a point of contact with a previous scholar’s writing, or an act of turf-claiming.[Pierre Bourdieu, writing about academic scholarship in general calls this phenomenon “citalogy”––the study of citational politics––and notes: “the strategic function of a reference may be apprehended in its modality: humble or sovereign, impeccably academic or sloppy, explicit or implicit and, in this case, unconscious, repressed [...] or knowingly dissimulated” @BourdieuMarketSymbolicGoods1993, 138] For literary scholars, as we’ve seen, recognition of sites of frequent (or infrequent quotation) can serve to display a scholar’s sense knowledge of where the field has collectively focused their attention.[Such patternings underpin what Edward Said called “contrapuntal” readings, where an individual scholar deliberately focuses on parts of the text others have neglected––for example, allusions to British imperialism in Jane Austen’s \*Mansfield Park\* @SaidCultureImperialism2012, 66.] Because the body of quotations in scholarly writing on a novel like \*Middlemarch\* produces a distribution of quotations, we are interested less in the sheer fact of a passage being quoted than in the particular conditions that contribute to that scholarly interest. While quantification is often thought of as entailing hyper-precision, our quantitative data on quotation patterns allows us to study reading and writing practices that individuals do not consciously experience in numerical terms. In addition, quantitative analysis allows us to study quotation patterns in a manner that is not about sharply-defined categories (some passages are “well-known”, others aren’t) but gradient (many different degrees of knownness are possible). As such, our study is quantitative but does not over-emphasize numerical precision as a means of interpretative method. You won’t see many actual numbers in the results that follow: patterns in quotation as we understand them are matters of relative proportions and densities.

# **From Quotations to Textual Atmospheres**

Collections of scholarly quotations form diffusive, aggregate objects––what we call “textual atmospheres.” In this article, we use the term textual atmospheres to refer to the critical afterlife of quoted material as produced through scholarly print culture. Individual scholars sense them both when reading (a passage read many times before in other scholarship will stand out from the rest, simply because of familiarity) and when writing (the history of previous quotations will shape which passages seem unavoidable, played-out or as yet under-appreciated). Textual atmospheres are an important part of the “critical afterlife” of a text (which also encompasses e.g. citations, allusions, adaptations). For the purposes of this article, we’ve confined ourselves to the community of scholars, but textual atmospheres can in principle emerge anywhere that texts are being excerpted and reproduced.[For a study of how selective textual reproduction reshaped ordinary Victorians’ understanding of British Romantic writers, see @MoleWhatVictoriansMade2017] Indeed, atmospheres exist even where texts are reproduced not in written but only in speech: every time a person alludes to a line from \*Hamlet\* or mentions the opening sentences of \*Moby-Dick\*, they increase the atmospheres around those specific portions of a much longer work. Similarly, even in the particular scholarly context we’re interested in––the post-war Anglophone academy––, atmospheres can emerge in other ways besides print: teaching is an obvious example, where the passages the teacher reads aloud, writes on a board or displays on a screen will acquire atmosphere for the students. That said, the methods we develop and the results we discuss in this article pertain to atmospheres produced through scholarly publications.

Our term “atmosphere” is itself a kind of quotation. We have taken it from some passing remarks by the Polish philosopher Roman Ingarden (1893-1970) in his 1932 book *\*The Literary Work of Art\**. There, he claimed that any given reader of a long literary work “is, from the beginning, under the influence of” a “literary atmosphere,” due to the smaller “moments” of the work which the broader culture has taken to be “characteristic.” By this Ingarden didn’t mean an immaterial Zeitgeist à la Hegel but something much more materially grounded. It is through “verbal or written transmission of moments [...] to another reader” that specific portions of a work become characteristic for a shared understanding within a given community. [@IngardenLiteraryWorkArt1973, 349. Ingarden’s “characteristic moments” resemble Matthew Arnold’s even more vaguely defined “touchstones” within the history of poetry. For a discussion of this term and its relationship to Victorian excerpting culture see @MenkeTouchstonesTitBitsExtracting2014, especially 564-565] In particular, Ingarden insisted that scholarly writings––“[a]ll “critical” articles, essays, discussions, attempts at interpretation, historico-literary studies, etc.”––contribute significantly to the development of atmospheres, “train[ing] the reader to understand the work in a certain way”, specific to the reader’s time, place and community.[@IngardenLiteraryWorkArt1973, 349. Ingarden’s atmosphere can be contrasted with another diffusive term from 1930s critical theory, Walter Benjamin’s aura, which understands reproduction as producing a sense of distance between original and reproduced text.]

Ingarden gives us a vocabulary for describing quotation conventions and the aggregate object they produce. For scholars in particular, atmospheres signal how the reproduction of text can produce particular disciplinary norms—something close to what Pierre Bourdieu has called the scholarly “habitus.”[@BourdieuHomoAcademicus1988] Throughout this article we use the term “textual” atmosphere, rather than Ingarden’s “literary” atmosphere, because it captures the specific medium through which this diffusive object emerges: texts. A primary text on the one hand, and a large collection of scholarly texts on the other. Indeed, textual atmosphere is not inherently “literary”: it need not be limited to literary studies and literary texts. One could study the textual atmosphere produced by philosophers around a frequently-quoted philosopher’s writings, or by sociologists around a key work of sociology. What distinguishes a concept like textual atmosphere from, say, Michel Foucault’s concept of discourse is that textual atmosphere are collections of statements that are empirically studied in particular sites: specific journals and their collections of quotations.[@FoucaultArchaeologyKnowledge1972, 27. See also part 2, “Discursive Formations,” where Foucault describes discourse as a “dispersion” of heterogeneous statements, which resembles our own conception of atmospheres.]

To think about the afterlife of texts in scholarly quotation as “atmospheres” is to understand the patterns we trace in them as appearing differently at different scales. [See Dora Zhang’s discussion of atmospheres as decentralized phenomena: @ZhangNotesAtmosphere2018] While Ingarden claimed that “a specific ‘trend of the times’ can be clearly distinguished [...] in a given period”, [@IngardenLiteraryWorkArt1973, 350] we’ve approached our study with skepticism about how a “trend of the times” become perceptible, and at what scale. In approaching literary history, we do not assume that literary scholars quote from primary texts in ways that are neatly coherent. Because literary history tends to operate on a smaller scale, the patterns and distribution of what parts of a given literary text are used over time, we lack a uniform methodology for describing literary historical phenomenon for scales between the very large and the very small.[See @JayHeyWhatBig2017 and other articles in this “Writ Large” special issue; see also @BuurmaSlowMetadata2020, @EnglishShiftingScalesLiterature2016 and @Saint-AmourMedialHumanitiesManifesto2019 as well as critiques of scaleability raised by @TanoukhiScaleWorldLiterature2008, @TsingNonscalabilityLivingWorld2012]

We focus on the “quotation” and on a collection of academic journals over six decades not in the service of f a bigger, faster version of disciplinary history, but because quantitative methods put pressure on some of the familiar categories of humanities study––things like authors, works, and texts––because they force us to reckon with the implicit units and scales of analysis in our work.[In restricting our attention to a small textual feature––the quotation––our work shares the spirit of productively “reductive” reading developed by @AllisonReductiveReadingSyntax2018]

When studied in aggregate, patterns in quotation in our corpus of scholarship, are at a mid-level scales, too large for humans to analyze qualitatively, too small for patterns to be really robust.[See @FrowMidlevelConcepts2010 for more on the utility of “mid-level” scales as an antidote to the usual black box of "the social" in the sociology of literature for conceptual work in literary history.]

# **The “Middlematch” Project: Uneven Textual Atmospheres of Quotation**

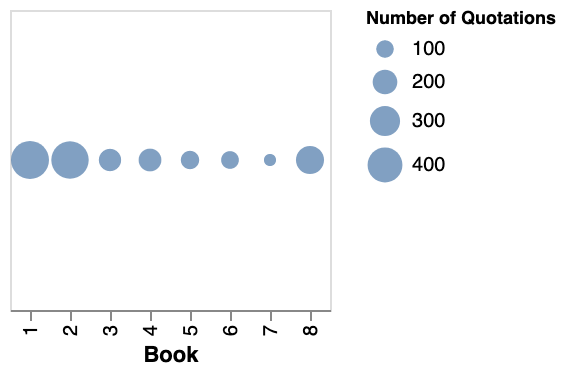
In “Middlematch,” our study of scholarly quotations from \*Middlemarch\* in the journal repository platform JSTOR, our algorithm identified 482 articles written between 1960 and 2015 that quoted some part of Eliot’s novel, totalling 1,794 quotations across 365 journals. We also identified the number and location of words quoted from the novel along with other metadata about that article: what journals these articles were published in, the years they were published, and the article’s title.

What patterns did we find in this data? First, that quotation was focused overwhelmingly, yet unevenly, on the first half of Eliot’s novel. The table below (Fig. 2) presents one view of the frequencies of quotations drawn from each of the novel’s 8 “Books”––the term for the novel’s original serial installments. From the table, you will see that Book 1 is the most frequently quoted book of the novel in our corpus, with about 27% of the total quotations, followed closely by Book 2, while Book 7 is the least frequently quoted, with just 49 quotations (or 3%). Endings are of course a popular place for scholarly discussion, and Book 8 is unsurprisingly the third most quoted, though at 14% it has almost half the frequency of Books 1 and 2. Overall, across the entire period we study, atmospheres have been thickest around the first few Books and the final one. Fig. 3 below, in which the size of each circle is proportional to the number of quotations from each Book, offers a visual representation of this pattern.

Fig. 2 : \*Middlemarch\* analysis: total quotations per Book

| **Book in \**Middlemarch\**** | **Number of quotations detected** |
| --- | --- |
| 1 | 475 |
| 2 | 460 |
| 3 | 164 |
| 4 | 171 |
| 5 | 111 |
| 6 | 104 |
| 7 | 49 |
| 8 | 260 |
| **TOTAL** | **1794** |

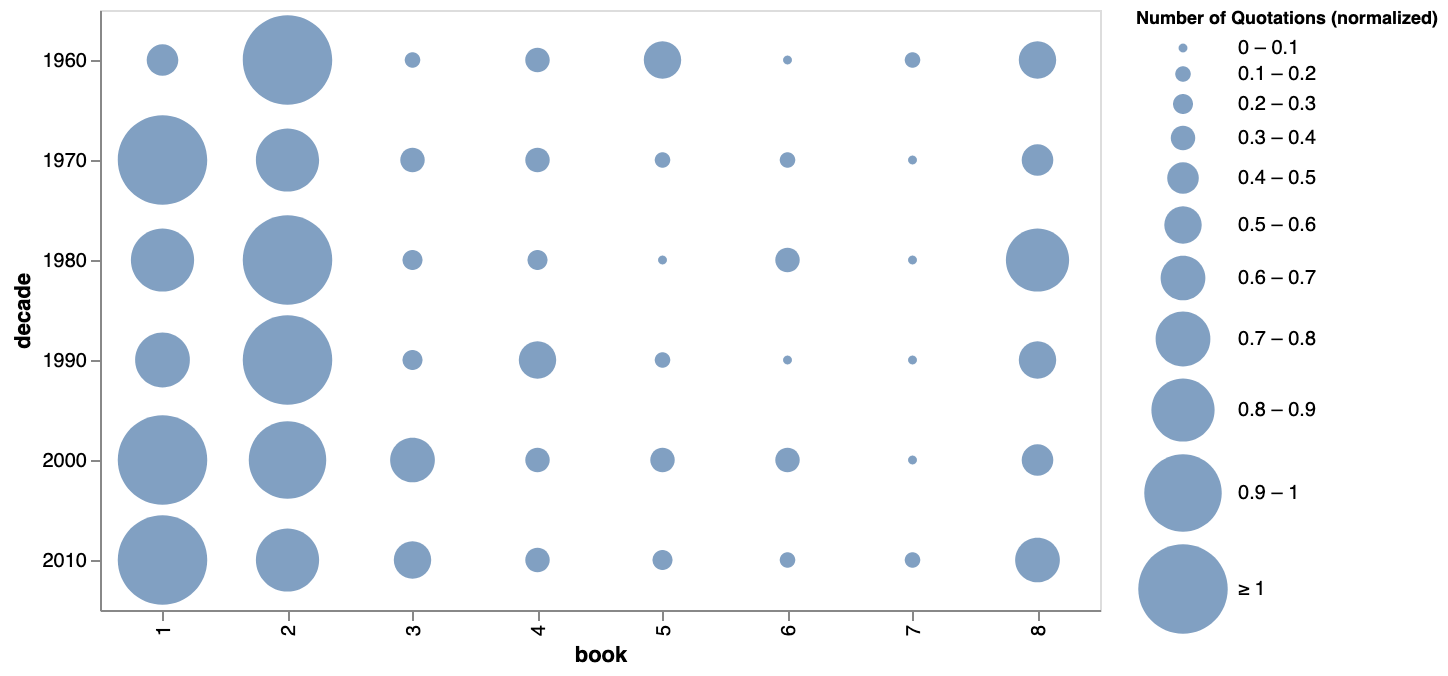
Fig. 3 : \*Middlemarch\* analysis: total quotations per Book



The table and visualization above display our initial findings synchronically: all of the results are grouped together according to Book. But as we’re interested in how these patterns of scholarly quotation change, or remain stable, over time, we also want to examine our quotations diachronically––to separate out the data on \*Middlemarch\* quotations by the publication date of each article. In graphs that follow, the horizontal axis indicates the textual unit (i.e. Book, chapter) of \*Middlemarch\*. The vertical axis indicates the decade of publication. As the legend to the right indicates, the size of the circles on the chart indicates how often passages from a given part of the novel were quoted in our corpus. The largest circles here represent the most frequently quoted parts of the novel in each decade, while the smaller circles represent little to no quotation detected in the given decade.

Fig. 4 below is a diachronic counterpart to Fig.3 above, and shows that the textual atmospheres of \*Middlemarch\* are not only uneven from Book to Book but also uneven across time. The column of large circles above “2” indicates that Book 2 has been consistently among the most quoted in every decade, and the column of tiny circles above “7” indicates Book 7 is among the least quoted. Nonetheless, all Books show some variation in quotation frequency over time, with Books 3-6 in particular varying across the middle range of frequencies.[^7] The differences between the synchronic and diachronic graphs shows that the scale of analysis determines what patterns we detect.

Fig. 4 : \**Middlemarch\** analysis: quotations per book, by decade

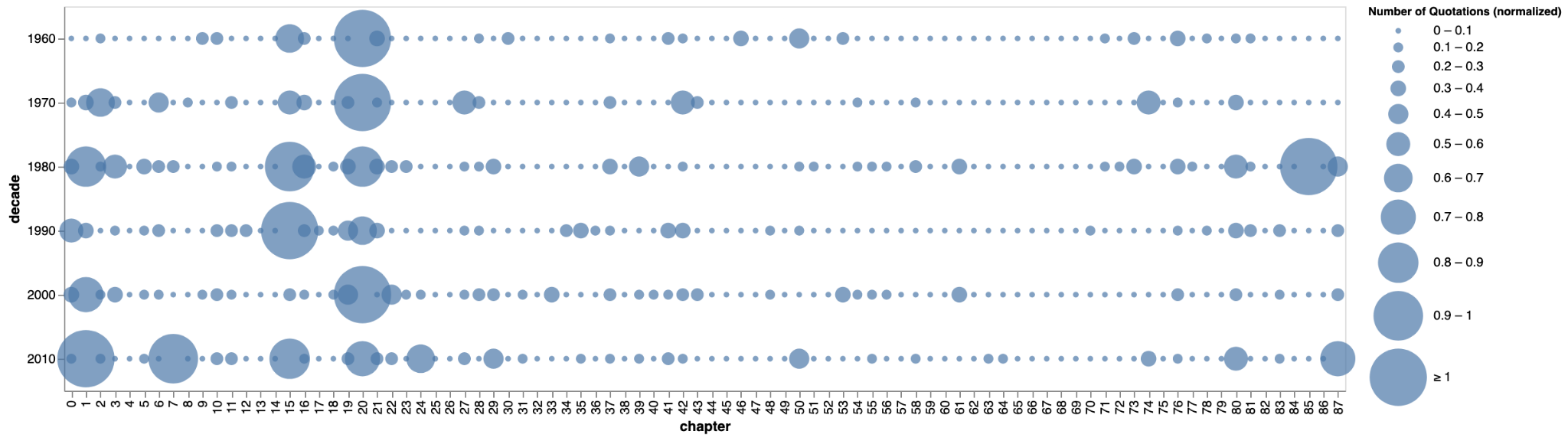


[^7]: Because scholarly publishing has grown significantly from 1960s to the 2010s, our corpus contains fewer articles from earlier decades and increasingly large numbers from more recent decades. In order to account for differing levels of scholarly publication in mid-twentieth century, we’ve opted to visualize the number of quotations proportional to the total number of quotations from \*Middlemarch\* in a given decade, rather than the raw count––a process known in statistics as “normalizing.” Each circle on the graph is sized according to the total quotations scaled so that the values fall between 0 and 1, with 1 (the largest circle), representing the most quoted book in that decade.

Shifting the scale of analysis proves similarly revealing when we examine quotation patterns not only at the level of the Book but also at the level of the chapter. \*Middlemarch\* has 86 numbered chapters, plus a “Prelude” chapter (labeled as chapter 0) and a “Finale” (labeled 87). The graph below (Fig. 5) shows the distribution of quotations per chapter across the entire corpus. The sizing has the same meaning as in the previous graph, with larger circles indicating more quotations from that part of the novel in the articles from the given decade. The most frequently quoted chapters are Chapter 20––the chapter which contains the “well known” squirrel’s heartbeat passage quoted in our epigraphs––and Chapter 15 (a chapter devoted to the doctor Lydgate’s research and backstory). Unlike Chapter 20, which has consistently medium- to large-sized circles across all decades, we see a gradual rise to prominence of Chapter 15 in the 1980s and 1990s (some of the largest circles for those decades, contrasting with only small or medium circles in other decades), which we investigate in more detail in section 4 of this article. The chapter-level view shows that not only has scholarly attention in this sixty-year period tended to favor the first half of the novel, and particularly the first two Books, it has been even more focused on a handful of particular chapters.

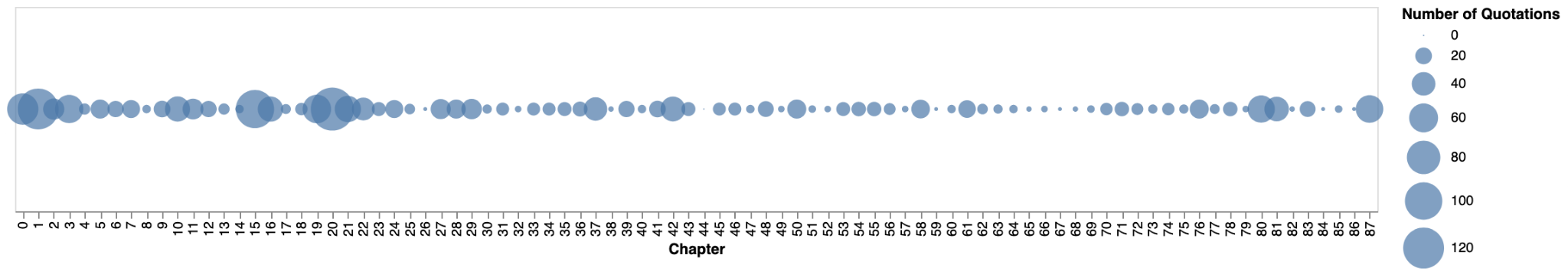
Some of the broader patterns already detectable in the graph of Book totals above continue to hold true here: although quotation frequency shifts among the early chapters, early chapters of the novel are among the most relatively frequently quoted in all decades, while chapters from the latter half (especially from Chapter 60-70) are consistently among the least quoted. On the other hand, the textual atmospheres towards the end of the novel look more varied when broken down by chapter: while Chapter 80 has medium-sized circles in most decades, quotation of other chapters is less consistent. While in the 1970s, 1990s and 2000s passages from the ending are infrequently quoted, in the 1980s and 2010s late chapters are among the most quoted overall.

Fig. 5: \**Middlemarch\** analysis: quotations per chapter, by decade



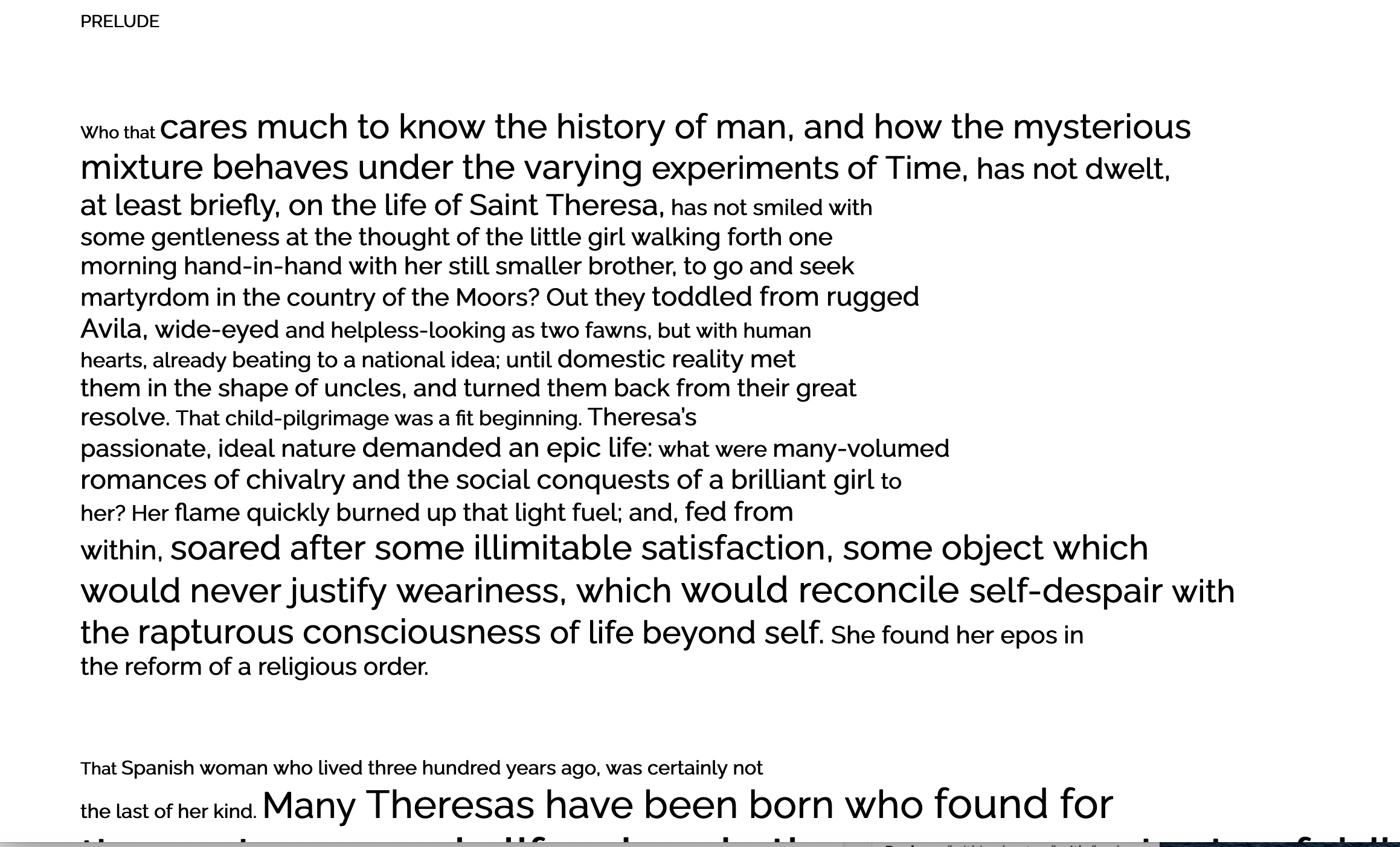
Even though we can see clear, and consistent patterns in what gets quoted that tend towards selected parts of the first half of the novel, nearly every single chapter gets quoted at least once during this 60-year time span, even if it is only quoted once during this entire time. This is easier to see in Fig. 6 below, a synchronic counterpart to Fig. 5. If you look closely, you can see there’s only one chapter, Chapter 44, in which our algorithm did not detect at least one quotation.

Fig. 6: \*Middlemarch\* analysis: quotations per chapter, visualized synchronically



Having observed some continuity and some discontinuity between the atmospheres at two scales of textual unit, we now move to an even more granular scale. At the level of individual paragraphs, we continue to detect unevenness in scholarly quotation (see Fig. 7). Like Fig.1, the image below is a screenshot of a browser we built that displays the text with font size proportional to how frequently it has been quoted in our corpus. Readers are invited to compare this passage from the opening of the novel with the squirrel’s heartbeat passage displayed in Fig.1. Both screenshots indicate that it’s not just Books and chapters that receive uneven scholarly attention––it’s individual paragraphs, and indeed sentences and shorter phrases within them. As Fig.1 shows, even within the “well-known” squirrel’s heartbeat passage, the first clause has only rarely been quoted within our corpus, and the first half of the paragraph still relatively infrequently. Only when the narrator moves into the grand generalizing sweep of “That element of tragedy” does the textual atmosphere become increasingly dense, only to thin out again after the squirrel’s heartbeat. Similarly, from \*Middlemarch\*’s opening paragraph, the first sentence and the second half of the penultimate sentence have been most frequently quoted, while the intervening sentences have been quoted less, but still occasionally.[To browse the text-level’s atmospheres beyond the screenshots displayed in this article, visit our Github repository: @Middlemarchcriticalhistories]

Fig 7: Screenshot of “Middlematch” Atmosphere Browser (Prelude)



These visualizations show the importance of understanding patterns in textual atmosphere at multiple scales. Each of these visualizations draws on the same data, but the unit of analysis––the Book level, the chapter level, or the paragraph level––show that it’s an even smaller portion of the novel that scholars have consistently quoted. It can simultaneously be true that Book 1 overall has the densest textual atmosphere, and that Chapter 20 from Book 2 has the densest atmosphere of any chapter. Indeed in the case of Chapter 20, it’s not the chapter as a whole but specific clauses within just a few sentences that account for this atmosphere. In other chapters (such as Chapter 15, discussed below) the chapter’s atmosphere is not attributable to a single isolated passage.

Even as the actual text that literary scholars have consistently quoted constitutes a much smaller portion of the novel than scholars may be comfortable admitting, our quantitative approach also shows that there are subtle patterns in what scholars quote. Scholars don’t quote in a vacuum. They quote––and they return to quote the passages other scholars quote––because this part of the novel has a history in scholarly discourse. At all scales––the synchronic, bird’s-eye view and the diachronic view at the decade by decade level, the book-level, chapter-level and paragraph-level––our results display highly selective atmospheres. By this we mean that among all the words of the novel that scholars might in theory quote, in practice their choices tend to cluster around a much smaller subset. Scholarly quotations, and the atmospheres they produce, are unevenly distributed, but in relatively consistent ways.

This uneven, selective distribution of scholarly quotation complicates questions of scholarly agency. Individually, a scholar may believe they’re constructing their own “reading” of a novel based on precisely those passages that best support their argument, but at these larger scales it seems as though some of these choices are primed by the existing scholarly environment, in other words by the textual atmosphere. While the fact of scholars responding to other scholars might not not seem like a surprising point––as scholars often respond to claims made by other scholars about a specific passage and build at least some of their arguments on a text’s “famous” passages––what we want to propose is that these activities constitute evidence of collective practices baked into literary scholarship. Viewed in aggregate, most scholars’ readings of \*Middlemarch\* are built on a foundation of the same few passages. While a literary scholar might look at these findings and come away with a dire warning about all the swaths of the novel that remain relegated to what Margaret Cohen calls “the great unread,” we argue instead that the passages scholars quote reflect their participation in a historically-specific scholarly community.[@CohenSentimentalEducationNovel1999, 23]

# **Whose \*Middlemarch\*?: Communities, Journals and the Situatedness of Scholarly Quotation**

What can a computer-assisted reading of quotations tell us about literary methods and history? Studying scholarly quotation this way can make visible the sorts of patterns that are harder to see: patterns of continuity, gradual change, or obscured histories of scholarly objects or attention. It’s only a slight exaggeration to say that today’s scholarly \*Middlemarch\* is not the \*Middlemarch\* of the 1970s. They may have Chapter 20 in common, but, as we’ve seen, even that canonical chapter has been less often quoted in the 2010s than it has been for most of the last 50 years. The \*Middlemarch\* that comes into view in excerpted quotations in twentieth- and twenty-first century scholarly articles takes a different shape with each decade in our corpus of journals, much like the Sappho’s poetry in the latter half of the nineteenth-century––though unlike Sappho’s poetry, in \*Middlemarch\* have an ‘ur-text’ that scholarly quotations might be compared to.[This argument about the Victorianness of the figure we call Sappho as a function of such practices of critical anthologizing comes from @PrinsVictorianSappho1999] The exact contours of these shapes––Middlemarch’s textual atmospheres––shift: at various scales of the novel, in synchronic and diachronic perspectives, and at the scale of individual journal itself as a particular institutional infrastructure.

Studying scholarly quotations with the help of quantitative methods allows us to see \*what\*, \*when\*, and \*how\* scholars quote at a scale larger than an individual author or article. This is important even for understanding acts of quotation that might appear to be quite individual. Consider, for instance, the anomaly of chapter 85 in literary scholarship. This chapter near the end of the novel is relatively infrequently quoted in mid-century scholarship. Our algorithm detected no quotations from this chapter in any of the articles from the 1960s or 1970s. But the chapter-level graph above shows that an atmosphere began to form around this chapter in the 1980s as the number of quotations intensify. How do we make sense of this change? In a 1980 issue of \*Nineteenth-Century Fiction\*, three prominent scholars––Barbara Hardy, J. Hillis Miller, and Richard Poirier––wrote a three-part essay in a special issue of the journal commenting on the style of this chapter. The journal quoted the entirety of Chapter 85 prior to their essays, which themselves were detailed close readings chock-full of quotations. In the articles themselves, there is an echo of Ingarden’s account of “characteristic” passages: J. Hillis Miller, for instance, claims that the chapter is a “typical example” of Eliot’s realism, one that “can stand by synecdoche for the whole.”[@HardyMiddlemarchChapter851980, 441] What the three scholars in 1980 identify as “typical” is, in their words, a stretch of text which is ordinary, unremarkable and undistinguished from most of the novel which surrounds it. While a single issue with three articles is responsible for the bulk of the quotations, this selection of a chapter that had not received extensive scholarly commentary to date needs to be understood in the shadow of the bulk of quotations from chapters in the first half of the novel. In selecting a “typical” example that was not one of the more frequently quoted passages, Hardy, Miller and Poirier make a move similar to Said’s contrapuntal reading: one shaped by the textual atmosphere of some of the more “well-known” quotations from the first half of the novel.

Patterns in what scholars quoted from Eliot’s novel across this six-decade-period are messier than something that could neatly be mapped on affiliation with a particular school of literary criticism of method. Take, for instance, Chapter 15. This is a chapter with some scholarly attention––3 articles quote from it in the 1960s, 7 in the 1970s. This chapter then becomes much more central in the 1980s and 1990s with 29 different articles quoting from this chapter––even more than the number quoting from chapter 20 (See Fig. 5), before dropping off again in the 2000s. When we look at the text of the quotations from this chapter, we had a hunch might contribute to the chapter’s focus: about 1/3 of the articles quote the narrator’s opening meta reflections about Henry Fielding and historiography while the rest quote mostly from four paragraphs describing the doctor Lydgate and his medical and scientific ambitions and from a few paragraphs at the end of the chapter in the melodramatic account of his relationship with a Parisian actress. Titles of the 26 articles quoting the chapter seem to suggest a New Historicist interest in the novel’s dialogue with intellectual history––“Middlemarch, Realism and the Birth of the Clinic” (1990), “An End to Converting Patients’ Stomachs Into Drug-shops: Lydgate's New Method of Charging His Patients in ‘Middlemarch’” (1998) or “The Language of Discovery: William Whewell and George Eliot'” (1998)––and with feminist readings of the novel: “The Strange Case of Monomania: Patriarchy in Literature, Murder in Middlemarch, Drowning in Daniel Deronda” (1988), “‘Middlemarch" and George Eliot's Female (Re) Vision of Shakespeare” (1991), and “Narrative Voice and the ‘Feminine’ Novelist: Dinah Mulock and George Eliot” (1992). Assuming that this chapter’s focus was likely a product of a New Historicist interest in the chapter’s engagement with nineteenth-century scientific discourse and with a 1990s wave of feminist scholarship focusing on the end, we had hypothesized that these scholarly trends would quote from different parts of the chapter.

To historicize a narrow period of literary scholarship, like the 1980s and 1990, means looking beyond monolithic descriptions of a method to what scholars actual produced. While we had set out to investigate instances where shifts in scholarly methods lead to distinct shifts in content: what we found was points of overlap: The fact that this chapter (and the handful of frequently quoted paragraphs in it) becomes a shared body of evidence for a methodologically eclectic range of scholarship. Both new historicist and feminist scholars did contribute to increased quotation from this chapter in these decades, the parts of the chapter they quoted did not straightforwardly map onto the chapter’s content. While some pieces––like four articles on historiography—quote only the opening paragraph of narrator’s meta-reflections. Articles like “Narrative Voice and the ‘Feminine’ Novelist: Dinah Mulock and George Eliot” quoted only from the first part of the chapter, focused not on women and the theater but on men and science. What might appear to be a sudden “rise” (and fall) in chapter 15’s quotation is, in fact, quotations from several different methodological schools.

The stories that we might tell about scholarly quotation practices depend heavily on the units of analysis we take in representing these quantitative findings (the book or chapter level, the sentence or paragraph level, synchronic or diachronic analysis) and the techniques of visualization. Each analysis of chapter 15 draws on the same data, but the unit of analysis––the book level, the chapter level, or the sentence and paragraph level––show us that it’s an even small portion of the novel scholars have consistently quoted. This finding is one of the most important, because it shows how this kind of method can tell us something not only about a particular novel and its scholarly history, but about the way that literary scholars use evidence.

What are quotation patterns evidence of? For one, they are evidence of institutional infrastructures. JSTOR articles do not exist in a vacuum: they come from particular journals, each of which offers a tangible site for studying the collective norms around what parts of the novel are quoted within that journal’s history––the textual atmospheres of that work. One of the reasons we chose JSTOR is because the repository includes full runs (through 2015) of major journals within literary studies 1950s and 1960s: genre-specific journals (like \*Novel\* or \*Studies in the Novel\*, both founded in the 1960s) or period-specific journals (like \*Victorian Studies\* or \*Nineteenth-Century Fiction\*, founded around the 1950s).

While up to this point, we’ve argued that there are roughly consistent patterns of uneven quotation from the novel, we see a different story at the level of individual journals, where the textual atmospheres vary more significantly. Take two very different journals–– \**George Eliot - George Henry Lewes Studies\** (henceforth *\*GEGHLS\**) founded in 1992 from an earlier specialist newsletter in Eliot studies, and this journal, *\*New Literary History\** (henceforth *\*NLH\**), founded in 1969. When examine patterns in what articles in these journals quote from *\*Middlemarch\** compared to the quotations from all other journals represented in the JSTOR corpus, we find several important differences. Articles in \*GEGHLS\* tended to focus more on minor character names: Casaubon, Garth, Ladislaw, Mary, Farebrother, Carp, Brooke.[The frequency of character names, relative to all other words in the text ––a combination of what is called term frequency and inverse document frequency (tf-idf)––is as follows: Casaubon is the most distinctive word in *\*GEGHLS\**, with a tf-idf of 0.001019; followed by Garth, ranked 0.001006, Ladislaw ranked, 0.000994; Mary, ranked 0.000868; Farebrother, 0.000667; Carp, ranked 0.000556; and Brooke, 0.000549. We derived this analysis using a method called “most distinct words,” inspired by work on the subject by @AllisonStyleScaleSentence2013. We here introduce a different method for studying what scholars quote beyond simply the location of passages. We offer it as a first foray into the content of \*Middlemarch\*’s quoted passages, which could also include analysis of parts of speech, dialogue vs narration, patterns in vocabulary. Interested readers can consult our GitHub repository @middlemarchcriticalhistories for further analysis of this kind.] Indeed, quotations from this journal seem to suggest a picture of disciplinary specialization itself, in which more obscure passages and characters are discussed more than those that are “well-known.”

By contrast, articles in \*NLH\* stand out not for what they quote, but for how little they quote.

\*Middlemarch\* has been cited in 25 articles in \*NLH\*, written between 1973 and 2013. While 25 articles have cited––that is, named––the novel \*Middlemarch\*, our algorithm was only able to detect 2 of the 25 as quoting text from the novel. First and foremost, then, \*Middlemarch\* exists for \*NLH\* as a point of reference rather than a text to be quoted. In the 2 articles that do quote from the novel, the quotations center on passages in the first half. In this respect, it reflects the general pattern we discussed at the start of section 3, where around 71% of all quotations came from the first half of the novel. Whereas the most distinctive words for *\*GEGHLS\** picked out various minor characters, \*NLH\*’s most distinctive words include “Celia”, “Dorothea” and “uncle” (i.e. Mr Brooke), all more major characters.[See @middlemarchcriticalhistories for more.] (As a final curiosity––no scholar in \*NLH\*’s history has quoted the squirrel’s heartbeat passage.) The fact that so many articles reference \*Middlemarch\* without quoting from it is not entirely unexpected expected from articles written by and for a readership where most scholars’ primary focus is not \*Middlemarch\*, George Eliot or even nineteenth-century British literature. We could speculate about the causes for \*NLH\*’s apparent lack of \*Middlemarch\* quotations: does the journals founding focus “the problems of literary history” translate into scholarly reading and writing practice that centers less on intensive parsing of individual passages close readings of an individual work?[@CohenNoteNewLiterary1969, 3.] What is most important about this glimpse of writing about \*Middlemarch\* is that it offers a glimpse of a tacit convention and technique (not large enough to be a methodological practice) that we see not just in single scholar, but in the journal’s writing in aggregate.

What our study of quotations in micro-communities of scholars writing about particular chapters or in the case of \* GEGHLS\* and \*NLH\* also reveals is the need reconceptualize the role of journals within narratives of change disciplinary history. What it helps reveal are the contours of Thomas Kuhn’s “normal science”: offering a disciplinary history paradigm shifts initiated by individual thinkers but on shared conventions which hold for a while, but are also subject to change.[@KuhnStructureScientificRevolutions2012] For us, those conventions are “the conventions and techniques that govern how scholars read and how they write,” as John Guillory put it, which are treating the corpus of scholarly writing, and the smaller subset of scholarly quotations.[@GuilloryHowScholarsRead2008, 11] We recognize can only make limited claims about or generalizations from a journal with only a few dozen articles over six decades. Without treating the body of JSTOR articles as a stand-in for scholarship as a whole, we show the that understanding some of these tacit convention starts by reckoning with everyday infrastructures of scholarship. Here, in focusing on quotation in the everyday practice of journals we follow Buurma and Laura Heffernan’s efforts to recover a fuller account of the history of the discipline, one that does not just take “the scholarship of professors working at a handful of elite universities as evidence of the main line of the discipline’s theories and practices.” [@BuurmaTeachingArchive2020, 2. Other scholars, like Elaine Auyoung, have acknowledged the presence of lay reading practices within the conventions of literary study, though Auyoung’s own readings take as a representative example just five scholars. See @AuyoungWhatWeMean2020.] As we’ve seen, changes in what parts of the novel are quoted are as much the produce of the preferences and idiosyncrasies of the parts of institutional and even scholarly––a kind of tacit house style––as they are a given decade’s scholarly trends. Literary scholars quote––and they return to quote the passages other scholars quote––because this part of the novel has a history in critical discourse. Sometimes the presence of this pre-history can lead us to not quote from a set of passage––as we see happening with quotations of Chapter 15 or the focus on minor \*Middlemarch\* characters in the specialist journal \*GEGHL\*––and there are ongoing chapters that critics seem to return to and quote again and again.

# **Conclusion: reading, knowledge, canons**

In this article, we’ve developed a method to study how patterns of scholarly quotation form historically shifting atmospheres around selected passages of a literary work. Studying quotations has allowed us to produce a disciplinary history whose primary agents aren’t individual scholars but shared conventions. Our findings for the atmospheres of \*Middlemarch\* have shown that data on textual atmospheres needs to be examined at multiple scales—the book, the chapter, the paragraph; the half-century, the decade; the journal, the journal issue—because each scale offers distinctive insights that are not necessarily detectable at other scales.

We have traced textual atmospheres in one specific historical and disciplinary frame: the practices of Anglophone literary scholars from the 1960s to 2015, in order to be able to study a fundamentally collective dimension to how the discipline operates. Scholars don’t quote in a vacuum. They decide what to quote in part based on certain passages’ place in the history of scholarly writings. Sometimes the presence of this pre-history can lead scholars to avoid quoting a passage, while there are other passages scholars return to and quote again and again. As such, quotations in scholarly writing operate as a site for empirically studying the collective products of disciplinary history. Through aggregation and quantitative analysis, quotations serve as concrete evidence of collective practice.

How might attending to the atmospheres produced through quotation reshape our understanding of the discipline? We want to leave readers with three possible directions.

First, all reading is sensitive to atmospheres. We, the authors of this article, both had our first reading of \*Middlemarch\* shaped by the atmosphere that exists around the squirrel’s heartbeat passage. As a teenager, [Author 2] repeatedly read the squirrel’s heartbeat passage on a poster about famous British writers in a classroom at school. [Author 1]’s first encounter with the squirrel’s heartbeat passage was as a quotation in an article assigned while reading another Eliot novel, \*The Mill on the Floss\*. When we first read \*Middlemarch\*, that passage stood out to each of us as a familiar place to linger and reflect. Textual atmospheres allow a small number of passages from novels to reach people earlier than the rest of the novel (indeed, for many it’ll be the only part of the novel they ever “read”). While our study has focused on scholarly readings, textual atmospheres can also shape non-scholarly readers’ encounter with a text. People don’t simply build up knowledge about a novel in a linear fashion as they read page by page. Instead, some passages of it arrive piecemeal before a reader opens the book, and some passages (often the same parts that arrived early) return after reading, sometimes many times, to reinforce their centrality in a reader’s memory. For all kinds of readers then, an encounter with a literary work is shaped by collective phenomena which a reader can’t control and in most cases can’t consciously recognize. Quantitative methods can reveal the collective conditions underpinning an aspect of the reading experience that can seem anecdotal, intuitive and idiosyncratic.

Second, quotation conventions shape disciplinary conditions of knowledge. Textual atmospheres show that, besides reading works of literature, literary scholars must also develop knowledge of a collectively authored anthology of key passages, quotations, and prior discussions of the works they study. This anthology is nowhere printed and never explicitly taught, and as such is always in flux. Quotations are material manifestations of that virtual anthology, and point to a discipline’s collective textual knowledge. Collectively, these quotations by many scholars, across time, create atmospheres that determine the conditions of knowledge for that work. When a portion of text has dense atmospheres, this facilitates ongoing knowledge production around it, as was the case for \*Middlemarch\*’s Chapters 15 and 85 in the 1980s, for example. For literary studies, whose paradigmatic object of study is a literary text, what parts of what texts get quoted are at the heart of the discipline’s knowledge production.

Finally, scholarly writing makes canons. In literary studies at least, the concept of canonicity has been strongly associated with titles of works and names of authors. But if we take canonicity to refer to anything that is widely read, discussed, analyzed, republished—in short, anything that is “well known”—then portions of text much smaller than an entire literary work also demand to be studied in terms of canonicity. Textual atmospheres reveal \*intra\*textual canonicity, i.e. the canon internal to a single text. Canonical works and canonical passages interact in complicated ways that our test case of \*Middlemarch\* can only begin to hint at: while \*Middlemarch\* as a title has been stably canonical for the entire period we’ve examined, its canonical passages have been more varied. As such, the \*Middlemarch\* that was canonical in the 1970s is not exactly the \*Middlemarch\* that’s canonical today (squirrel’s heartbeat aside). This may seem paradoxical, but perhaps this is how all canonical works sustain their place, offering enough different passages to allow for fresh takes on a well-known work.

Canons are more than just the mere presence or absence of titles on lists, then. Our test case of \*Middlemarch\* is both widely cited \*and\* widely quoted, but atmospheres are equally important when considering under-represented authors and texts within the canon. It’s not enough to cite more names and titles on syllabi and bibliographies if in practice this leads to only a few isolated passages–or even no passages at all–gaining denser atmospheres through quotation. Instead, scholars must grapple with the fact that their everyday practices of quotation make and remake each text as it moves through history.