

Gender Dynamics and Critical Reception: A Study of Early 20th-century Book Reviews from *The New York Times*

Matthew J. Lavin^a

^aUniversity of Pittsburgh, Pennsylvania

ARTICLE INFO

Peer-Reviewed By: Barbara Green, Matthew Wilkens
Article DOI: 10.22148/001c.11831
Dataverse DOI: 10.7910/DVN/C10VIQ
Journal ISSN: 2371-4549

ABSTRACT

This paper focuses on book reviews at the turn-of-the century United States in order to underline fundamental compatibilities between large-scale, computational methods and book historical approaches. It analyzes a dataset of approximately 2,800 book reviews published in *The New York Times* between January 1, 1905 and December 31, 1925. Several machine learning scenarios are employed to investigate how the underlying reviews constructed gendered norms for reading and readership. Logistic regression models are trained and tested to evaluate how effectively lemma frequencies predict the perceived or presumed gender of an author under review. The paper discusses four different feature selection scenarios, as follows: (1) No terms removed, (2) Stop words removed, (3) Stop words, gender nouns, and titles removed, and (4) Stop words, gender nouns, titles, and common forenames removed. For each scenario, the top lemma coefficients are discussed and interpreted. Tracing the norms (gendered and gendering) of *The New York Times Book Review* in the early twentieth century demonstrates that even the summary-driven book reviews played an important role in mediating hierarchies of taste and distinction. Further, the paper seeks to demonstrate that cultural analytics methods can be used to investigate a range of research questions related to authorship, publishing, circulation, and reception.

Now is a moment when a consumer can locate a review of Aristotle's *Poetics* as easily as a review of *Black Panther*, an Echo Dot, or the Grand Canyon. As a result, it is perhaps easier than ever before to lose track of the review as a complex genre with rich historical underpinnings. In fact, numerous distinct prototypes converged over time to develop into the genre we recognize today. The *Oxford English Dictionary* traces the usage of the term *review* meaning "an account or critical appraisal of a book or (now also) a play, film, concert, etc." to pamphlets as far back as 1649. Lucien Febvre and Henri-Jean Martin write of "a whole series of bibliographical journals" in the 17th century, followed by the *Journal des Savants* (January 1665), which they describe as the first known periodical to include a

“review of recent publications.”¹ Frank Donoghue describes the rise of the book review during the 18th century against the backdrop of “a kind of limbo” between “an age of substantial aristocratic support and the fully developed literary market of the nineteenth century.”² Michael Gavin, similarly, suggests that review magazines “regularized and familiarized” criticism in the mid-18th century.³ Though associated with the broader category of literary criticism, book reviews emerged as a distinct subset of reader responses. In the nineteenth century, Joan Shelley Rubin argues, large-scale changes in both production and distribution of books and periodicals “partially enhanced the stature of the genteel critic,” but this growth in status was relatively short-lived as, in the first decades of the twentieth century, genteel critics’ centrality began to decline.⁴ Janice Radway describes a generalized “anxiety about the potential destruction of traditional culture in the wake of the transformations effected by rapid social change.”⁵ Set against this backdrop, book reviews at the turn of the twentieth century emerged as one site among many where contentious debates about cultural capital were taking place.⁶

For the most part, there is consensus around these broad historical points, but much of what we think we know about book reviews is based on close analysis of specific examples, which have been elevated for one reason or another. Patrick Collier calls for a better understanding of “what the object of knowledge *is* in modern periodical studies” in order to avoid “a plethora of micro-studies that have incommensurate aims and methods, are not speaking to each other, and thus are not contributing to an overall understanding of how periodicals functioned within the cultural field at the turn of the twentieth century, or of that cultural field itself.”⁷ Case studies have a crucial role to play in cultural studies but, when taken in isolation, they are especially vulnerable to what Andrew Piper calls “the problem of generalization,” or the challenge of “how to move from part to whole.”⁸ Precisely how did the changing norms of cultural valuation and prestige function for book reviews? How did book reviews fit into a larger context of triangulations among cultural producers and consumers? How were reviews of a particular work or author shaped by the broader readership and reception landscapes associated with their work? What effect did categorical norms have on the ways authors and texts were categorized, assessed, and circulated? The large scale, computational methods of the burgeoning cultural analytics subfield can help address questions like these.

What follows is an analysis of book reviews published in *The New York Times* between January 1, 1905 and December 31, 1925. I have constructed a dataset of approximately 2,800 documents for use with several machine learning scenarios to investigate how the underlying reviews construct gendered norms for reading and readership. I have taken up the question of how gender—a crucial categorical norm—affected how authors and texts were described and evaluated. Tracing the norms (gendered and gendering) of *The New York Times Book Review* in the early twentieth century demonstrates that even the summary-driven book reviews—“a collection of book reports to consumers on the readability of new titles” according to Richard Kluger—played an important role in mediating hierarchies of taste and distinction.⁹ I also hope to demonstrate, by example, how cultural analytics methods can be used with documents like historical book reviews to investigate a range of similar research questions.¹⁰

Scholarship on Turn-of-the-Century Culture and Readership

For readers of *Cultural Analytics*, my work will appear most overtly in dialogue with Ted Underwood, David Bamman, and Sabrina Lee’s “The Transformation of Gender in English-Language Fiction.”¹¹ Responding to “both the gender positions ascribed to authors as biographical personages, and the signs of gender they used in producing characters,” Underwood, Bamman, and Lee argue that an ostensibly paradoxical shift has taken place between 1800 and the present day—first, that “gender divisions between characters have become less sharply marked over the last 170 years” and second, a “decline in the proportion of fiction actually written by women, which drops by half (from roughly 50% of titles to roughly 25%) as we move from 1850 to 1950.”¹² Underwood, Bamman, and Lee’s large scale study is provocative and convincing. It does not engage, however, with book reception or the broader landscape of paratexts—authorial and allographic—that most certainly mediated the gender norms of English-language fiction between 1800 and the present day.¹³ Meanwhile, scholars in a periodical studies context have written extensively about how distant reading can be best employed to analyze a range of content published in periodicals. This body of scholarship is generally enthusiastic about the potential of

such intersections, and it raises several important notes of caution for future work.¹⁴ Applying supervised machine learning directly to large-scale corpora of book reviews strikes me as an appealing area for continued convergence between cultural analytics and periodical studies.¹⁵

In addition to speaking to some of the preoccupations of periodical studies, my research question—how did *New York Times* book reviews between 1905 and 1925 describe published work in relation to perceived gender lines—revisits three dominant preoccupations of previous scholarship on readership at the beginning of the twentieth century: the exchange of economic and symbolic capital in a “field of cultural production”; the disruption of cultural hierarchy in the construction of early 20th-century taste; and the historical feminization of particular cultural ideals.¹⁶

First, Pierre Bourdieu’s work on economic and symbolic capital attempts to explain how authors, publishers, critics, readers, etc. engaged in symbolic struggles that, in aggregate, shaped a “field of stylistic possibles” or “space of possibles.”¹⁷ Bourdieu’s *Distinction: A Social Critique of the Judgement of Taste, The Logic of Practice, and Rules of Art: Genesis and Structure of the Literary Field* are concerned with issues that relate, directly or indirectly, to this subject. James English’s *The Economy of Prestige*, Loren Glass’s *Authors, Inc.*, and Mark McGurl’s *The Program Era*, among others, demonstrate the applicability of Bourdieu’s framings to a U.S. context. Allison et. al.’s “Quantitative Formalism: An Experiment,” Piper’s “How Cultural Capital Works,” and Jurafsky et. al.’s “Linguistic Markers of Status in Food Culture: Bourdieu’s *Distinction* in a Menu Corpus” exemplify an ongoing ambition in digital humanities to interrogate and better understand the apparent cultural mechanisms and structures of cultural valuation that Bourdieu initially described.

My work also revisits the broader concerns of a distinct cluster of studies of readership that have engaged with historicizing the parameters of cultural hierarchy at the turn-of-the-century, mostly in the United States and the United Kingdom. Although many articles and books in this field have engaged with Bourdieu’s core ideas, “middlebrow studies” has largely focused on the concept of *middlebrow* as a referential keyword and a foregrounding concern.¹⁸ The term was not coined until 1925, and the degree to which the concept predated the term is a much larger debate than I could address here. Initially, this scholarship was situated in relation to the

historic origin of the terms *highbrow* and *lowbrow* but the field has steadily widened. As Cecilia Konchar Farr and Tom Perrin have argued, the field has grown to the point that “we no longer feel, as we once did, that a gloss on the term *middlebrow* is a vital component of any piece of writing on it.”¹⁹ My own work is not concerned with the keyword *middlebrow* but does intersect with issues of cultural hierarchy raised by others in this field. Two foundational works of middlebrow studies, in particular, are crucial to the mapping of cultural hierarchy at the start of the twentieth century: Radway’s *A Feeling for Books: The Book-of-the-Month Club, Literary Taste, and Middle-Class Desire* and Rubin’s *The Making of Middlebrow Culture*.

Like others, Radway’s primary concern is whether cultural spaces between high and low were permissible, and how the “values associated with one form of cultural production were wed to forms and values usually connected with another.”²⁰ She argues that a preponderance of “how-and-what-to-read literature” in the late nineteenth century argued the virtues of “reading as a goal-directed activity” and positioned itself against cheap fiction, as well as “the sensual, somatic pleasures of the body” associated with reading for enjoyment.²¹ Notions of middlebrow, she argues, emerged from this utilitarian context. The “scandal of middlebrow,” Radway writes, “was a function of its failure to maintain the fences of cordoning off culture from commerce.”²² Although her work does not engage directly with the book review as a form, it remains one of the most important and often cited touchstones on the idea of middlebrow, as well as the pre-1925 logic of reading that no doubt influenced *The New York Times Book Review* at the start of the twentieth century.

Rubin, in contrast, is directly concerned with the function of book reviews in the early 20th-century United States. In this sense, Rubin’s work contributes to a larger body of scholarship predating “distant reading” that studies book reviews quantitatively or systematically.²³ Rubin describes middlebrow as a series of cultural mechanisms by which genteel values of the nineteenth century “survived and prospered, albeit in chastened and redirected form, throughout the 1920s, 1930s, and 1940s.”²⁴ According to Rubin, the “news” approach to book reviews was “virtually the only mode of presentation in the daily press during the antebellum period.”²⁵ These reviewers tended to allocate as much if not more column inches to book summaries as they did to their evaluations. Their reviews often treated amusement and edification as equally valid reasons for praise. In turn, Rubin argues, genteel

editors of monthlies and quarterlies were critical of the news approach to book reviews for developing an overly close relationship with book publishers.²⁶ Book reviews' primary exigencies—summarization and evaluation of published work—represented an enticing space for many competing cultural principles to interact.

My analysis of book reviews in *The New York Times* engages with how cultural taste was gendered at the turn of the twentieth century. Previous scholarship has demonstrated convincingly that the rise of the mass market was deeply intertwined with rapidly changing and ardently contested notions of manliness, masculinity, womanhood, and femininity. Much of this work has tied feminization to the concept of middlebrow, but it is widely argued that such feminization began well before the term *middlebrow* was ever used. According to Jaime Harker, “Depending on the context, ‘middlebrow’ can mean ‘middle class,’ ‘effeminate,’ ‘polluted by commerce,’ ‘mediocre,’ or ‘sentimental.’”²⁷ Beth Driscoll’s *The New Literary Middlebrow: Tastemakers and Reading in the Twenty-First Century* locates gender as one of eight interconnected characteristics “through the middlebrow of the twentieth century to the new literary middlebrow of the twenty-first century.”²⁸ Driscoll argues that middlebrow culture tends to be both “female and feminized,” as it is “implicated in a wider pattern of gender discrimination that runs through the literary field.”²⁹ Radway, similarly, has argued that “the debate over books and reading was a heavily gendered debate in the sense that cultural conservatives always associated the threat of cheap fiction and passive reading with the dangers of ‘aimless,’ ‘indolent,’ and ‘ardent’ femininity.”³⁰ Reading habits and other patterns of cultural consumption were reconfigured along gender lines during the early years of the twentieth century.

Dataset: Book Reviews and Metadata

As Maria DiCenzo has argued, even in a distant reading context, “reading periodicals (closely or deeply) for their discursive and visual content … remains central to research engaged in expanding historical and cultural fields.”³¹ The same, I and others would argue, should apply for literary studies approaches to subjects like genre and canon formation. To this end, I have endeavored to work with *The New York Times*’ TimesMachine’s front end, and my underlying dataset, in equal

proportion. *The New York Times Book Review* was founded in 1896 as an eight-page Saturday supplement. Prior to this, book reviews appeared in *The New York Times Sunday Magazine Supplement* and in columns like “New Publications” in the news section of the daily paper.³² Some reviews or announcements of prominent books continued to appear in the news section even after 1896. Where the news section of *The New York Times* was a seven-column broadsheet in 1896 and an eight-column broadsheet after April 1, 1913, the book review supplement was a three-, four-, or five-column tabloid, depending on the year. The Saturday supplement moved to Sunday in 1911 and, between 1920 and 1922, it was titled *The New York Times Book Review and Magazine*. After 1922, it became an increasingly respected component of the Sunday edition of the *New York Times*, as many as 96 pages in length at its height.³³ From the outset, *The New York Times Book Review* followed previous established norms by including publishing statements before review material, with information such as the book title, reviewed author, reviewed author’s other books, book format, publication place, publisher, number of pages, and book price. Typographical markers like drop cap initials and book decorations were also common accompaniments to content.³⁴

The supplement consistently included reviews and non-review material, such as features about well-known authors, coverage of prominent book auctions, letters from readers, and literary gossip. Reviewed works were presented in several different ways, including one review of many books, recurring features like “Latest Works of Fiction” that clustered reviews together but offset each book with its own publishing statement, and single-author, single-book reviews, a minority of which were signed by well-known reviewers. Artwork occasionally appeared with a review, and author photographs became increasingly common over time, especially for highlighted authors. By the 1920s, artwork was integral to *The New York Times Book Review*’s page design.³⁵

The book reviews for my analysis are drawn from a larger set of approximately 27,000 pieces of content published in *The New York Times Book Review* between January 1, 1905 and December 31, 1925. In addition to covering *The New York Times Book Review* during crucial years of development, this date range targets the period when the news approach to book reviews, according to Rubin, was the dominant norm. I downloaded metadata for these articles using *The New York Times*

Article Search API and used contextual clues (such as a \$ in the headline and page number BR1-BR25) to generate an initial list of potential single-work reviews. I then culled this list by setting aside non-review content, very short reviews, reviews comprised primarily of quoted material from the reviewed work, and multi-book reviews. Multi-book reviews can be further split into two distinct groups: a single review of many books, or a scanned pdf that is actually a cluster of several reviews, treated as one unit either deliberately or inadvertently.³⁶ For each review, I hand-coded a gender label to describe the assumed gender of the author of the reviewed work. A review that referred to an author as “he” or “Mr. Smith” was labeled “assumed male,” and a review calling the reviewed work’s author “she” or “Mrs. Smith” was labeled “assumed female.” Throughout this article, I have used the phrases “male-labeled-reviews” and “female-labeled-reviews” to describe these groups. If a reviewed work had multiple authors perceived as a mix of genders, it was labeled “multi” and, in the less frequent cases when a reviewer avoided the language of gender altogether, it was labeled “unknown.”³⁷ However, multi-labeled reviews and unknown-labeled-reviews were removed from the training and test sets for the purposes of this article.

In the interests of procedural transparency, I want to describe some pertinent details about how review text was processed in preparation for machine learning. Johanna Drucker has argued, “Designing a text-analysis program is necessarily an interpretative act, not a mechanical one, even if running the program becomes mechanistic.”³⁸ Although many of Drucker’s generalizations are subject to debate, the interplay of rule-based text processing techniques and interpretive judgments that she describes is an integral feature of quantitative hermeneutics. For this study, optical character recognition (OCR) text was used for all computations.³⁹ A natural language processing library was used to remove punctuation, tokenize, and lemmatize.⁴⁰ Lemmatization, which combines inflections like *charm* and *charms* into one token (called a lemma), is not thought to be necessary in text classification and may in some cases even reduce classification accuracy slightly, but it can make term coefficient results more readable by reducing what essentially seem to be repeat terms. To remove as many OCR errors from the model as possible, I also employed automated spellchecking.⁴¹ Relative lemma frequencies were weighted by inverse document frequency (TF-IDF).⁴²

Three limitations to the scope of this study should be noted. First, I do not claim that the reviews analyzed in this study represent broader trends in book review language. Rather, I hope that a larger sample of reviews (with comparable metadata) from various periodicals will be developed. Second, I deliberately limit my discussion of the identities of book reviewers in this article. Forming conclusions on this subject would also require a larger historical sample, as most reviews in *The New York Times* at this time were published without crediting a review author. Finally, I do not wish to make claims about the “objective” or “ground truth” of the genders or sexes of authors discussed in these reviews, nor do I endorse the idea that gender must be described in such essentializing terms. My goal here is to trace evidence of gendering in review language, not to use reviews to predict gender for information gathering purposes.

In their introduction to a special issue of *Feminist Modernist Studies* on “Feminist Modernist Digital Humanities,” Amanda Golden and Cassandra Laity argue, “Until recently, DH has been prominently associated with scientific neutrality, ‘big data,’ quantification and the ensuing practices of distant reading or macroanalysis.”⁴³ Although I think this charge against digital humanities is overstated, I share the concern that “purportedly ‘objective’ knowledge systems can and do inscribe exclusionary, hierarchical assumptions.”⁴⁴ My work resists the assertion that cultural analytics methods necessitate binary thinking, overconfidence, or disavowal of their own limitations. At the same time, I embrace Catherine D’Ignazio and Lauren Klein’s perspective that “the double-edged sword of data shows just how important it is to understand how structures of power and privilege operate in the world.”⁴⁵ Following and building on prior scholarship by Matt Jockers and Gabi Kirilloff; Eve Kraicer and Andrew Piper; Susan Brown and Laura Mandell; Richard Jean So, Hoyt Long, and Yuancheng Zhu; and Underwood, Bamman, and Lee, my work seeks to adopt a binary, temporarily, as a way to interrogate it.⁴⁶ Cultural analytics can do more to interrogate and historicize categories like gender, as well as trace how these categories were socially constructed.

Methods: Logistic Regression and Feature Coefficients

Regularized logistic regression is a well-established machine learning method that many digital humanities or cultural analytics practitioners have employed.⁴⁷ Texts (in this case book reviews) are divided into a train and test set. Reviews in the training set are used to build a machine learning model (logistic regression) and a series of predictions for the book reviews in the test set are generated. The overall performance of the model is then evaluated by comparing each prediction to its corresponding hand-coded gender label. For these tests, I included only “assumed male” and “assumed female” reviews in the training and test sets, as logistic regression is designed for binary classification. Running a logistic regression with a small dataset of gender-tagged book reviews can immediately answer two questions:

1. Using book review language as machine learning features, how well can a logistic regression model predict the presumed gender of a book’s author?
2. Which features are the most “useful” (i.e., provide the most information) in making these predictions?

The first of these questions is addressed by running predictions and evaluating the results, though several factors can complicate how we interpret a model’s overall performance. First, an unequal presence of male and female labels must be considered. The book reviews dataset I assembled contains 2,173 male-labeled-reviews and 715 female-labeled-reviews, a ratio of approximately 3 to 1. This count makes for a striking portrait of gender imbalances, and it likely downplays gender disparities in *The New York Times* at this time. This ratio was only achieved after female-labeled-reviews were oversampled to ensure at least 20 samples for each represented year. For my start and end years (1905 and 1925), I conducted a full inventory of single-work book reviews and found ratios closer to 6:1.⁴⁸

To establish stable performance baselines, I have retrained each model using hold out cross-validation with resampling of 1,000 different randomized training set and test set membership combinations. Unlike a k-fold cross-validation or leave-one-out cross-validation, such resampling can represent performance metrics, such as precision, recall, and F1 scores, using measures of central tendency against a backdrop of a normally distributed range of results.⁴⁹ (Figures 1a and 1b demonstrate

such aggregated F1 scores.) To obtain still more information about the factors affecting performance, as well as the categories of language that best predict gender, I re-examined my research question using four feature selection scenarios (Table 1).

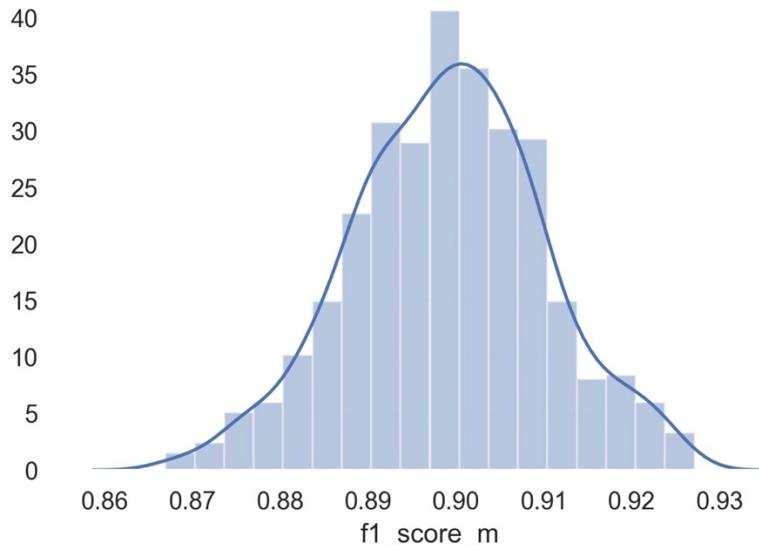


Figure 1a. Distribution of F1 Scores (1,000 Reshuffles) for Male Labels in Scenario 1 (Mean: 89.91; Standard Deviation: 0.011)

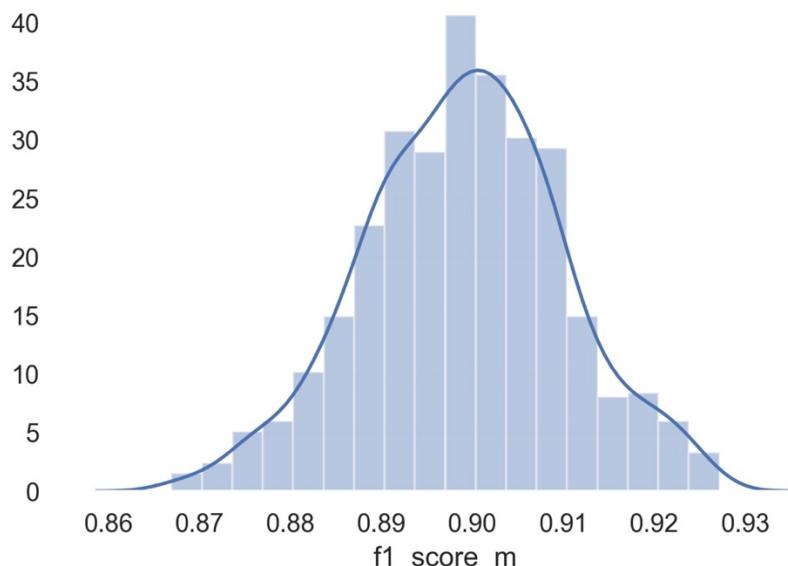


Figure 1b. Distribution of F1 Scores (1,000 Reshuffles) for Female Labels in Scenario 1 (Mean: 77.58; Standard Deviation: 0.025)

Scenario 1	No terms removed
Scenario 2	Stop words removed
Scenario 3	Stop words, gender nouns, and titles removed
Scenario 4	Stop words, gender nouns, titles, and common forenames removed

Table 1. Machine Learning Feature Selection Scenarios

Logistic regression is a preferred machine learning model for a study like this one partly because the rules used to make predictions are transparent. A model trained on term frequencies to predict whether a book reviewer regards the book's author as male or female generates numerical equations that score the likelihood of maleness and femaleness using the assumption that some terms' frequencies will correlate with this prediction. A model's coefficient scores provide an indication of how much importance each training feature was found to have, although how much a particular feature's coefficient score affects a given prediction is a more complicated question.⁵⁰

With this study, I have taken the additional step of aggregating coefficient scores over all training set and set test reshuffles. A feature appearing on the top of the aggregated coefficient list is the feature with the greatest mean coefficient, and the relative stability of that feature can be described using measures of central tendency. (Figure 2 demonstrates this stability with the example of the term *her*.) Throughout this paper, I will discuss lemma features that repeatedly and consistently informed gender label predictions.

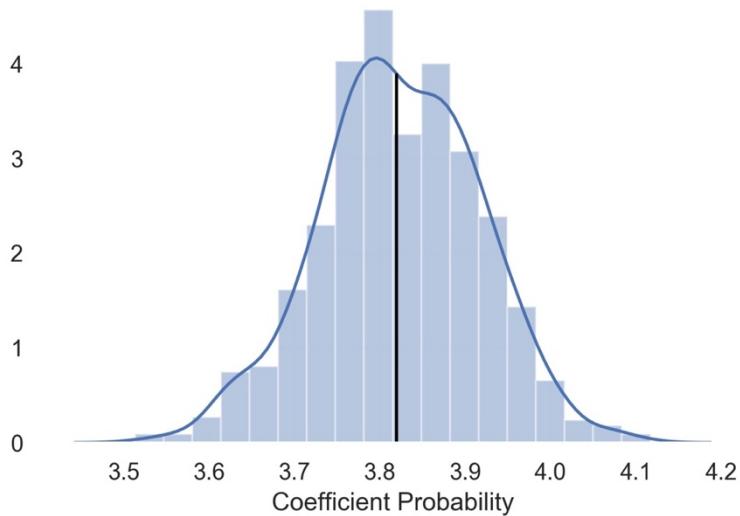


Figure 2. Distribution of Coefficient Probabilities for "her" (1,000 reshuffles) for Scenario 1 (Mean: 3.82; Standard Deviation: 0.094)

As the “Results” section of this article will show, an initial inspection of term coefficients suggests that surface level terms—e.g., pronouns, gendered nouns, gendered titles, and forenames—are strong predictors of gender. In my view, these seemingly innocuous terms are crucial to understanding how book reviews engage in gender making; additionally, machine learning can help isolate terms associated with the content of the books reviewed and the reviewer’s evaluation of a book, which is more difficult for a human reader to do consistently. In my attempts to find features that scholars focusing on the gendering of cultural taste might deem qualitatively meaningful, I focused on isolating and removing categories of features with overt gendered content.

Results: Model Performance and Term Coefficients

The Scenario 1 model considering all lemma features from the underlying dataset of book reviews can be configured to predict the assumed gender of the reviewed work’s author with accuracy rates between 78% and 90% and a mean accuracy of 86.1% (0.015 std). Recall and precision rates were higher for male-labeled-reviews even though two steps were taken to keep the model as balanced as possible. Table 2 summarizes performance measures for all four feature set scenarios. Performance is benchmarked with F1, precision, and recall scores for male and female labels, as well as overall accuracy measures. The model uses class weighting to maintain approximate parity between female and male recall rates despite the fact that male-labeled-reviews outnumbered female-labeled-reviews.

Scenario	Features Removed	F1 (F)	Precision (F)	Recall (F)	F1 (M)	Precision (M)	Recall (M)	Accuracy
1	None	0.776	0.773	0.781	0.899	0.901	0.897	86.10%
2	Stop words	0.712	0.687	0.743	0.863	0.881	0.848	81.50%
3	Stop words, gender nouns, and gendered titles	0.593	0.551	0.649	0.793	0.829	0.762	72.60%
4	Stop words, gender nouns, gendered titles, and forenames	0.581	0.538	0.638	0.786	0.823	0.754	71.70%

Table 2. F1, Precision, Recall, and Accuracy Scores by Gender for Scenarios 1-4

As one might expect, the predictive strength of the model decreases as directly gendered terms are removed (in steps over Scenarios 2, 3, and 4). At the same time, the regression generates a more robust predictor of male labels than female labels, even after class balancing is imposed to keep the labels relatively similar to one another. In fact, as gender pronouns, gendered titles, and common forenames are removed, precision and recall rates decline more steeply for female-labeled-reviews than for male-labeled-reviews.⁵¹ These performance metrics establish an interpretive baseline. The model has more male-labeled-reviews than female-labeled-reviews to work with, so it comes as no shock that categorical predictions for each label differ; yet, if the model could find no underlying categorical patterns, no amount of additional data would improve the model's predictive accuracy. In this model, more data leads to better performance, which is a general sign of robustness.

Coefficient scores for each scenario suggest several patterns. For Scenario 1, the top coefficients associated with female-labeled-reviews are terms like *her*, *she*, *mrs*, *miss* and *woman*. The information provided by categories of words is at least partially hierarchical: stop words have the highest coefficients (removed in Scenario 2), followed by gendered nouns and titles (removed in Scenario 3), and then forenames (removed in Scenario 4). After all these categories have been removed, *child*, *story*, *novel*, *heroine*, and *home* have the highest coefficient scores. For male labels, *his*, *mr*, *the*, *of*, and *that* are the top coefficients before any categories of features are held out (Scenario 1). Male-gendered nouns (*boy*, *man*) and forenames do not inform predictions of male labels, though honorifics like *dr*, *prof*, and *professor* are associated with male-labeled-reviews. By Scenario 4, *dr*, *british*, *volume*, *essay*, and *prof* have the highest coefficient scores, though some of these top terms are noticeable as early as Scenario 2. Tables 3a-3d list the top five term coefficients predicting male and female labels for each machine learning scenario.

Female Label Coefficients		Male Label Coefficients	
Term	Score	Term	Score
Her	4.78	His	2.07
She	3.82	Mr	1.73
Mrs	2.12	The	1.08
Miss	1.51	Of	0.78
Woman	1.23	That	0.49

Table 3a. Top Coefficients for Scenario 1

Female Label Coefficients		Male Label Coefficients	
Term	Score	Term	Score
Mrs	3.32	Professor	0.74
Miss	2.28	Dr	0.63
Woman	2.1	Prof	0.54
Girl	0.94	British	0.54
Mother	0.92	Volume	0.51

Table 3b. Top Coefficients for Scenario 2

Female Label Coefficients		Male Label Coefficients	
Term	Score	Term	Score
Child	1.1	Dr	0.65
Story	0.98	British	0.61
Novel	0.88	Prof	0.57
Heroine	0.84	Volume	0.56
Love	0.82	Essay	0.53

Table 3c. Top Coefficients for Scenario 3

Female Label Coefficients		Male Label Coefficients	
Term	Score	Term	Score
Child	1.19	Dr	0.67
Story	1.06	British	0.64
Novel	0.96	Volume	0.61
Heroine	0.95	Essay	0.58
Home	0.81	Prof	0.52

Table 3d. Top Coefficients for Scenario 4

Interpretation of Results

In this section, I want to raise three aspects of these results that connect to gendering in book reviews, as well as mediating hierarchies of taste. First, I want to return to Scenario 1 (which includes gendered personal pronouns, common forenames, and gendered titles like “Mr.” and “Mrs.”) to discuss terms that would otherwise seem mundane. Second, I examine the trends of reviewed books that emerge from these results, with particular attention to book subject matter that seems to split most clearly by gender. Such bifurcation reinforced a notion of discrete gender spheres at a time when much if not most literature by women was working to complicate or

blur such separations. Third, I advance the idea that a distinct set of gendering features related to authorship, genre, and process can be extrapolated from these results, and that these terms connect gendering to the work of cultural mediation.

Stop Words, Pronouns, Honorifics, and Forenames as Predictors of Gender

As much of this essay already suggests, book reviews have specific exigencies that may restrict how gender is portrayed, including norms for how the plot and the author of the reviewed work are discussed. A book reviewer encounters a book attributed to a proper name and makes inferences based on prior knowledge, guesswork, or some mixture of the two. The rules of inference are historically dependent; reviewers most likely believed they could determine authors’ “true” genders by making normative inferences. As Barbara Hochman argues, nineteenth century custom was supportive of “reading for the author.” Direct address (e.g. “Dear reader”) was common in fiction, and authors’ identities were often assumed based on the subjects they wrote about. Over time, however, “Amid growing uncertainty about how to conceptualize an author’s relation to book and reader, many novelists imagined the act of reading itself as a hostile attempt to ‘get at’ the withheld figure concealed behind the words on the page.”⁵² A known persona may come to mind when an author’s proper name is invoked, which would no doubt inform a reviewer’s assumptions; yet, inevitably, authorial figures were described in relation to a broader ideas about categories, including but not limited to masculinity and femininity.

Turning to the coefficients for Scenario 1, it may seem inconsequential or obvious that an author’s gender can be predicted by examining the frequency of terms like *her, she, mrs, miss, woman, his, mr, the, of, and that* in a book review. I included such terms in the first scenario, however, to demonstrate that deceptively simple terms can convey gender in unexpected ways. Notably, in this set of models, female gendered pronouns have higher coefficient scores than male gendered pronouns. Male-labeled-reviews, in contrast, are more readily associated with non-gendered function words. The linguistic norms of turn-of-the-century book reviews echo what de Beauvoir observed, that “being a man is not a particularity” to men— “[woman] determines and differentiates herself in relation to man, and he does not in relation to her; she is the inessential in front of the essential.”⁵³ It is possible that these results

are affected by the authorship signal of book reviewers, since, as Koppel, Argamon, and Shimoni have noted, male and female authors tend to use certain high frequency function words in predictably different ways.⁵⁴ However, the available metadata for my study suggests that reviewers between 1905 and 1925 were mostly men. This was true for reviews of books by assumed men and books by assumed women.⁵⁵ As a result, it seems unlikely that male and female reviewers writing in different ways would account for these results. Directly gendering female pronouns and gendered titles set in contrast to ostensibly non-gendered pronouns and titles seem to have contributed to a set of rhetorical framings that sanctioned men and women as different kinds of authors.

These patterns are especially meaningful when we consider their presence in texts like book reviews. The interplay of ostensibly obvious words frames what Bourdieu describes as the habitus and hexis of gender. For Bourdieu, habitus is a “set of dispositions which generates practices and perceptions” learned as early as childhood, which creates a “second nature” that shapes behavior.⁵⁶ Hexis is a specific, bodily instantiation of this habituation: “what is ‘learned by body’ is not something that one has, like knowledge that can be brandished, but something that one is.”⁵⁷ A category like gender is inscribed, bodily, beginning at a young age, and is “inseparable from a relation to language and to time.”⁵⁸ Gender norms, like norms in other categories, are established in what Bourdieu would call a “space of possibles” that textual representations—along with other constructors of social meaning like individual, bodily acts and broader sets of group behavior—are continually reinforcing and revising over time.⁵⁹ Book reviews, especially those taking the so-called news approach, present themselves as banal guides to particular texts but contribute to the habitus with every iteration.

Interpreting Indications of Book Subject Matter

The regression results from Scenario 4 seem to suggest a relationship between assumed gender and the subject matter of reviewed books. Before I proceed with discussing how these trends may have played out, I want to emphasize that drawing interpretive conclusions from term or lemma lists can be especially problematic. As various schools of lexical and semantic linguistics (before and after the poststructuralist turn) have maintained, individual terms acquire meaning in the

context of an intricate network of term-topic associations.⁶⁰ When coefficient lists are viewed, associations come to mind, but some of these associations are more complex than they might first appear. Contextual differences such as part-of-speech, word sense, and metaphorical usage have been conflated into one label, and lemmatization adds yet another level of reduction in this case. The hazards of forming conclusions from bag-of-words models are perhaps best articulated by Ben Schmidt, who cautions that the popularity of topic modelling is based on “a set of assumptions that are only partially true”—first, that co-occurring terms, “will therefore have a number of things in common” and second, that “if a topic appears at the same rate in two different types of documents, it means essentially the same thing in both.”⁶¹ Like Schmidt, I would argue that these navigational complexities do not preclude deriving interpretive conclusions from bag-of-words patterns. Rather, term co-occurrence “neither can nor should be studied independently of a deep engagement in the actual word counts that build them.”⁶² In the context of a book review, it may be tempting to assume a word describes an author, a book, a character in the book, or something else, but lemmas are potentially composed of descriptions at many or even all of these levels.

Exploring all coefficients comprehensively becomes impractical even at a scale of a few thousand short documents, since each lemma can appear in hundreds of reviews and that lemma may appear in many sentences in one review. Using the regression settings from Scenario 4, I hand-divided coefficients into several recognizable “first pass” groupings (Tables 4 and 5). I intend for these labels to provide a birds-eye view of a much more complex constellation of term usage in individual reviews. I grant that a different person might use different labels, or create any number of subgroups, and I do not mean to suggest that the words I grouped together co-occur in particular reviews. There might be many reviews of books where government is discussed and, quite separately, many reviews of books with some mention of the law. In turn, a word like *fact* or *essay* could be closely associated with terms like *law* or *history*. I present these groupings as heuristic constructions shaped by my subjective judgment, but also informed by examining the reviews directly.

Category	Lemmas (alphabetized)
Author and Text	Book, character, fiction, heroine, letter, novel, picture, plot, story
Domestic and Social	Background, child, custom, family, home, house, household, queen, social, society, widow
Feminized Virtues	Charming, clever, dear, feminine, friend, life, little, lovely, young
Marriage and Courtship	Handsome, heart, marriage, married, meet, romance
Time	Day, year
Economic Class	Labor, poor
Other	Come, color, know, lace, moor, real, type, vivid, want, wear, Wharton

Table 4. Interpretive Groupings for Female-Leaning Coefficients (Top 50 Lemmas)

Category	Lemmas (alphabetized)
Author and Text	Drama, essay, paper, play, translation, volume
Conquest and War	Army, battle, captain, empire, general, military, sea, ship
Geographic Locales	British, Germany, Japanese
Scholarship and Science	College, dr, prof, science, scientific, Shakespeare, theory, university
Government and Policy	Constitution, government, law, nation, national, policy, political, president, state, united
Procedural	Fact, history, lie, opinion, present, second, series
Uncategorized	Business, economic, fr, ill, jame, mile, sir

Table 5. Interpretive Groupings for Male-Leaning Coefficients (Top 50 Lemmas)

The observed term-category patterns depicted in Tables 4 and 5, while certainly informed by complexities of usage, are stable across many different training and test partitions and suggestive of potential connections. To contextualize some of my label choices, I derived statistics for each lemma, including: gendered document frequency of each lemma, term frequencies for each word contributing to the lemma's frequency grouped by part-of-speech tag, and the number of total synsets in Wordnet associated with each word/part-of-speech combination.⁶³ For example, the lemma *child* is found in 700 documents in the corpus; 36% of all female-labeled-reviews; and 24% of all male-labeled-reviews. The words *child*, *children*, and *childs* are lemmatized to *child*, and the terms combined have four part-of-speech variants, as listed below (Table 6).⁶⁴

Word	Part of Speech	Frequency
Child	Noun	622
Child	Verb	1
Children	Noun	792
Childs	Noun	1

Table 6. Words and Part-of-Speech Counts for the Lemma 'child'

In this example, three of the four terms are associated with the same four synsets:

1. child.n.01, "a young person of either sex"
2. child.n.02, "a human offspring (son or daughter) of any age"
3. child.n.03, "an immature childish person"
4. child.n.04, "a member of a clan or tribe"

The combination of term, part-of-speech, and synset data for the lemma *child* suggests an especially stable set of terms and potential uses. All but one root term seems to be used as a noun. There are only four synsets, and all four synsets are closely related. I placed it in the category "Domestic and Social" because of its association with families, but it could also be associated with biographies, children's literature, or (at least in theory) a derisive statement about an author.

In contrast to *child*, the lemma *play* appears in 681 documents, 25% of male-labeled reviews, and 20% of female-labeled-reviews. The lemma represents *play*, *plays*, *playing*, and *played*, and its part-of-speech variants are listed below (Table 7).

Word	Part of Speech	Frequency
Play	Noun	494
Played	Verb	263
Play	Verb	263
Plays	Noun	230
Playing	Verb	92
Plays	Verb	83

Table 7. Words and Part-of-Speech Counts for the Lemma 'play'

Further, these forms of the lemma *play* are associated with 52 synsets, 35 nouns and 17 verbs. Compared to most of the coefficients, *play* represents a broad range of potential uses. At the same time, the terms *play* and *plays*, tagged as nouns, are combined more frequently than the variants tagged as verbs, and a quick look at review headlines shows results, like “Shakespeare’s Poems,” “Bernard Shaw’s Latest,” “Seven Plays by Americans,” and “Drama Victorian and Modern,” so I placed the lemma under the label “Author and Text,” despite its clear complexities in usage.⁶⁵

In the regression results, female-labeled-reviews use terms more readily associated with domestic and social settings, marriage and courtship, and historically feminized qualities like charm and loveliness. The lemma *charming*, in this model, represents only the word *charming* (and not *charm*, *charmed*, or *charmer*), and it appears in 190 reviews. 214 uses are tagged as adjectives, and only nine are tagged as nouns. The lemma has one primary usage and meaning, yet there is still room for complexity. As one review of Dorothy Canfield Fisher’s *Understood Betsy* suggests, “There are some charming pictures of the simple, wholesome country life.”⁶⁶ This reference to *charming* refers to Canfield’s pictures of country life. In contrast, a review of Julia Ward Howe’s biography, written by her daughters, states, “When [Howe] and her two sisters grew up, so lovely and charming were they that they were known as ‘The Three Graces of Bond Street.’”⁶⁷ This use of the adjective *charming* refers to Howe’s personality. The presence of a lemma like *charming* could come from the review’s description of the author’s writing (as with Fisher), the biographical subject (as with Howe), or some other aspect of a book altogether. In the broadest possible sense, all of these uses are more likely to appear in reviews of books by those perceived as women in the aggregate, and that larger pattern is one example of what would remain invisible without cultural analytics methods.

A gloss of the lemmas I have grouped under “domestic and social,” “feminized virtues,” and “marriage and courtship” suggests the entanglement of gendered subject matter and gendering descriptions. The individual lemmas for *child*, *heroine*, *home*, *marriage*, *family*, and *young* are all in the top ten coefficients. It could be the case that lemmas associated with reviewed books’ subject matter are more likely to coincide with gender, but this is difficult to say because there is no static, objective line between a subject matter lemma and a descriptive lemma. Male-labeled-reviews

seem to favor lemmas associated with demonstrations of power and prestige, but that broad association seems to include subject matter and something more. The lemma *state*, which I have grouped under “Government and Policy,” is associated with the noun *state*, as well as entities like *the United States* and *Secretary of State*. The lemma appears in 989 reviews (37% of male-labeled-reviews and about 26% of female-labeled-reviews). It is associated with seven word/part-of-speech pairings, yet the nouns *state* and *states* represent 2,126 occurrences, whereas the verbs *state*, *states*, *stated* and *stating* combine to represent 241 occurrences. As a review of Sir Alfred Lyall’s authorized biography of Lord Dufferin (governor general of Canada from 1872 to 1878) describes, “There was a strong element of the Canadian government almost fanatically loyal to Great Britain, but multitudes looked for independence or union with the United States as the natural destiny of the Dominion. … It is not too much to say that the personality of Dufferin was instrumental in bringing about a change of sentiment and opinion.”⁶⁸ This text demonstrates how various references to multiple uses of *state* might intertwine, along with related terms like *statesman* and *statecraft*, both of which are mentioned in the review. More generally, the lemmas for *british*, *state*, *nation*, *political*, *law*, *government*, and *president* are all among the top twelve coefficients, which suggests in the most general sense that these likely subject matter keywords are more frequent among male-labeled-reviews.

I want to emphasize the potential significance of these results to literary studies in particular. Many traditional literary studies approaches to feminism at the turn of the twentieth century have suggested that the decades between 1900 and 1930 were crucial in breaking down the doctrine of separate spheres.⁶⁹ Other computational scholars have found word-level predictors for authors’ or characters’ gender in fiction that seem to echo and confirm these findings, but such studies have almost exclusively focused on 19th-century bifurcations. Matthew L. Jockers’s work on gender and authorial identity in *Macroanalysis*; Jockers and David Mimno’s “Significant Themes in 19th-century Literature”; and Jockers and Kirilloff’s “Understanding Gender and Character Agency in the 19th Century Novel” all discuss “the valorization of passive, domestic female behavior” in fiction, but they restrict their analysis to texts published before 1900.⁷⁰ Underwood, Bamman, and Lee, whose work traces character descriptions in a corpus that covers 1800-1980, observe that, “gendering of privacy and interiority was linked to a broader division

between public and domestic spaces.”⁷¹ However, term frequencies separating character gender such as *mind*, *spirit*, *passion*, *chamber*, *country* all converge by 1900; and terms such as *heart* and *room* begin with demonstrable associations to the feminine and become less associated with femininity. Broadly, they conclude, “it would appear that genres themselves were becoming less strongly gendered.”⁷² Reviews in *The New York Times*, in direct contrast to all of these articles, tell a story of women writing more fiction than men, and the doctrine of separate spheres alive and thriving. Piper and So argue that reviews in *The New York Times Book Review* between 2000 and 2016 “essentially reproduced the public/private split bequeathed to us from the nineteenth century”; perhaps this split has remained consistent in book reviews for more than 100 years.⁷³

I do not mean to suggest that Bamman, Jockers, Kirilloff, Lee, Mimno, and Underwood are incorrect in what they report. Rather, this large set of book reviews covers a 20-year period and probably distorts the trends one is likely to find by looking directly at a corpus of novels from the same time period. What *The New York Times Book Review* reviewed or opted not to review, along with how their reviews tended to describe books, are probable factors in this distortion, as if literary history has been reflected in a fun house mirror. Nevertheless, this example can remind practitioners in cultural analytics, and in literary studies, that direct examination of published fiction and nonfiction and examination of authorial and allographic paratexts can suggest very different interpretations. This points to the potential shortfalls of analyzing either one without considering the other, and it speaks to the importance of modeling authorship, production, circulation, and reception when doing the work of literary history.

Interpreting Authorship, Genre, and Process Terms

Presumably, the differences in gender norms that I have observed between reviews and novels from the same time period are a result of some combination of trends pertaining to book selection, book summarization, and book evaluation. For one, my dataset includes reviews of fiction and nonfiction, so the gendering of subject matter could be more pronounced in nonfiction, which would inform the domestic-romantic and war-scholarship-government bifurcations I have discussed. Further, newspapers may have been more likely to publish bestsellers and, as Underwood, Bamman, and Lee note, 31% to 42% of such bestsellers were by women between 1900 and 1930 (with a peak of 42% in 1930 before a long, slow descent).

More hand encoding by genre might shed some light on these questions, but Scenario 4 regression results provide an indication that genre affects book review gendering, since male and female authors are associated with distinct sets of authorial or publishing industry terms. Under the title “Author and Text,” the lemmas *book*, *character*, *fiction*, *heroine*, *letter*, *novel*, *picture*, *plot*, and *story* bend the model’s prediction toward female-labeled-reviews. These terms suggest references to storytelling strategies, especially in the context of fiction. “Author and Text” lemmas associated with female-labeled-reviews have higher coefficient scores, on average, than any of the apparent subject matter groupings. Term divisions for *fiction* and *novel*, in particular, suggest that *The New York Times* was reviewing more fiction by women than fiction by men, and/or more nonfiction by men than nonfiction by women. When compared with the Scenario 1 lemma counts for *she* (Figure 3), it might be more difficult to estimate the cut-off for *fiction* or *novel* (Figures 4 and 5) that would best split the data into “mostly male” and “mostly female” groupings.⁷⁴ Yet the regression model suggests that, as the counts for both of these terms increase, so does the probability of the review having a female label.

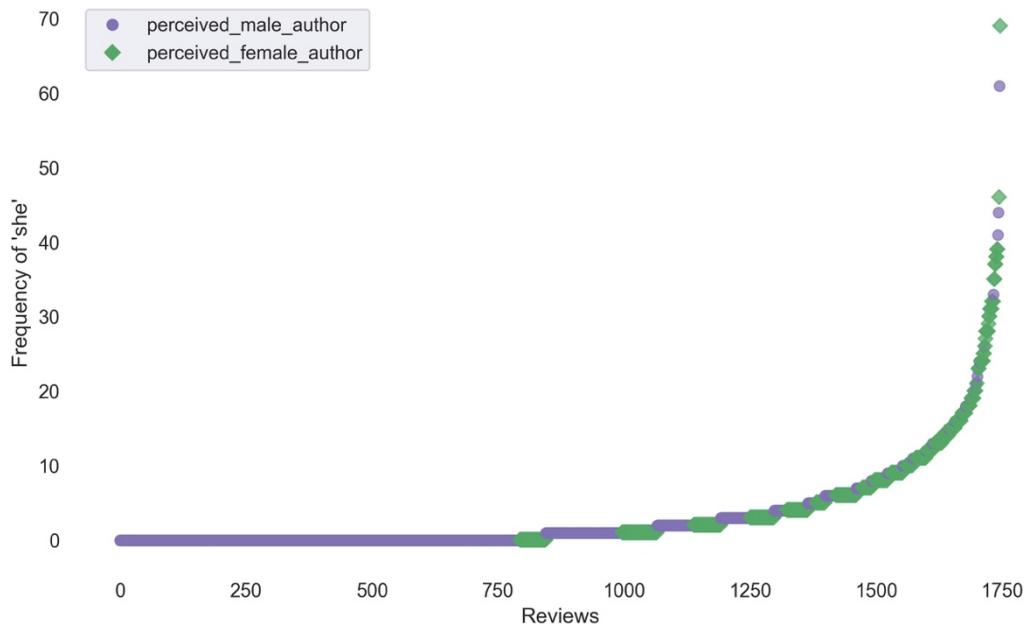


Figure 3. Scatter Plot of Counts for 'she' Colored by Gender Label

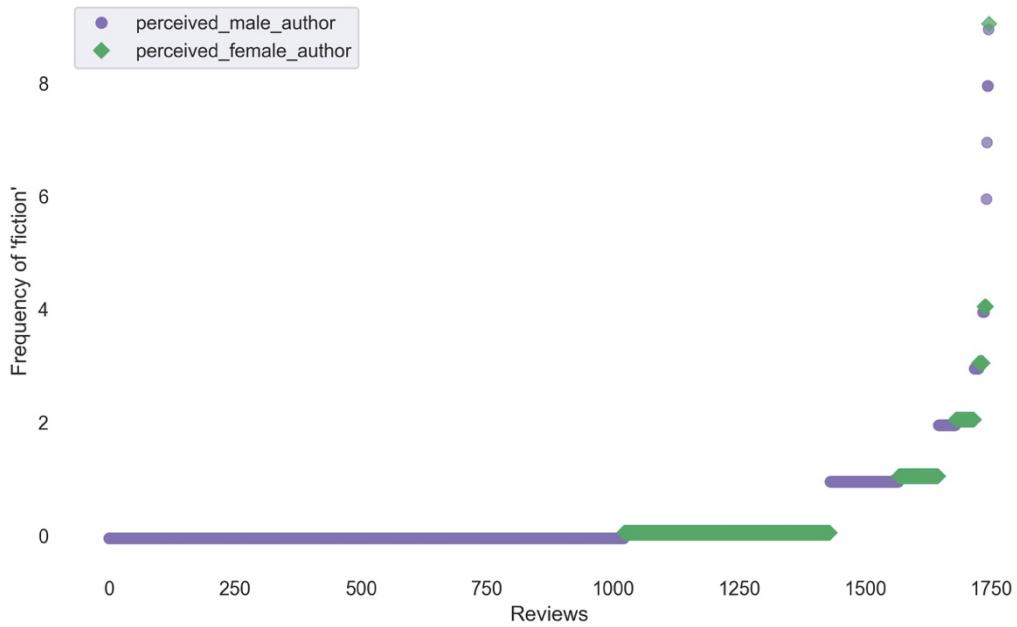


Figure 4. Scatter Plot of Counts for 'fiction' Colored by Gender Label

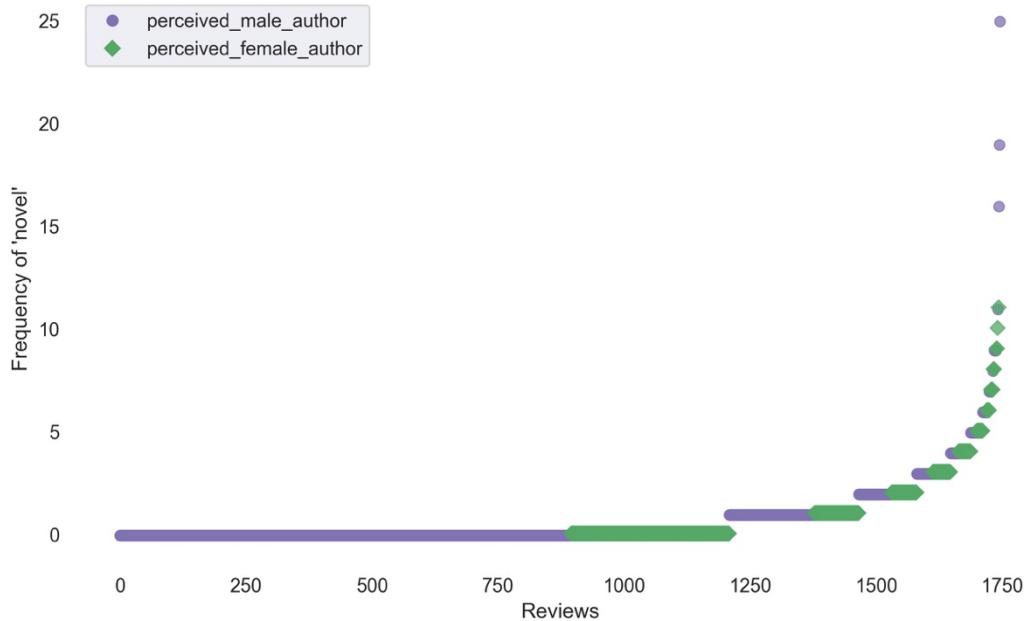


Figure 5. Scatter Plot of Counts for ‘novel’ Colored by Gender Label

For male-labeled-reviews, lemmas such as *drama*, *essay*, *paper*, *play*, *translation*, and *volume* have high coefficient scores. Coefficient scores for “Author and Text” lemmas associated with male-labeled-reviews are comparable to any of the subject matter groupings I have already discussed. In my qualitative categories, I have also added a category called “Procedural,” which includes the lemmas *fact*, *history*, *lie*, *opinion*, *present*, *second*, and *series*. Each of these lemmas could, to some degree, raise an image of a reviewer discussing an author’s general approach or specific choices. Many of these lemmas suggest associations with nonfiction, or a reviewer’s response to it, and this generalization is consistent with the fact that female-labeled-reviews do not seem to have a counterpart category. Taken together, these results provide reason to think that male-labeled-reviews are more closely linked to drama, nonfiction, recurring series, and gatekeeper functions, as well as multi-text series and new editions of prior work.

Largely because of the preponderance of “Procedural” terms, overall, regression coefficients for male-labeled-reviews appear more abstract than female-labeled-review coefficients. In other words, even in largely summary-driven reviews, we can observe patterns that suggest the gendering of concreteness and abstraction, with

male-labeled-reviews retaining a kind of privilege over female-labeled-reviews. In these reviews, authors perceived as male may be more often granted rhetorical space to disappear into their ideas and opinions. In a book review context specifically, this kind of privilege seems immediately relevant to the curatorial and mediating role that male authors and critics were so often granted. Read against this context, lemmas related to scholarship, academia, science, history, and economics may further suggest a male-dominated symbiosis among higher education, publishing, and book review apparatus. A larger sample could say even more about this pattern, especially if reviews from additional periodicals were included, but these results alone point to clear norms for male and female authorship across two decades of book reviews in *The New York Times*.

Concluding Remarks on Middlebrow Culture

I began this article by expressing two goals for this analysis of book reviews published in *The New York Times* between 1905 and 1925:

- (1) To demonstrate that even this newspaper's summary-driven book reviews played an important role in gendering reviewed authors, partly by gendering subject matter and genre norms, and partly by gendering the work of mediating taste and distinction.
- (2) To show the potential of using large-scale, corpus-based analysis of historical book reviews and other paratexts to revisit additional cultural analytics research questions.

To accomplish the first of these goals, I have focused on how machine learning classification can point to gendered patterns in both subject matter and structural vocabularies of book reviews. Assumed female authorship is associated with lemmas that suggest domestic settings, romance and marriage plots, and a constellation of feminized values. These differences of subject matter, in particular, suggest a culture of non-authorial paratexts mediating and reinforcing a sense of division between "what men write about" and "what women write about" that has not been observed when primary texts such as novels were analyzed using large scale, computational methods. Assumed male authorship is associated with lemmas that may imply subject matter like the military, government, academia, and status. Female-labeled-reviews are likely to have a greater number of lemmas associated

with fiction, whereas male-labeled-reviews are likely to have a greater number of lemmas associated with essays, plays, and series. Finally, lemmas reminiscent of cultural curation and remediation are associated with male-labeled-reviews. According to Driscoll, middlebrow suggests a specific constellation of values: “The literary middlebrow is middle-class, reverential towards high culture, and commercial; it is feminized, emotional, recreational, mediated, and earnest.”⁷⁵ Part of Driscoll’s point is that, over time, the various aspects of what we now call middlebrow became associated with one another, such that we would expect to see the work of mediation becoming feminized, but this effect does not seem present in this particular corpus.

Going further, terms associated with commercialization do not appear in the subject matter and structural coefficient lists for the machine learning scenarios above (Tables 4 and 5). Some overtly commercial terms may not have been used very often in book reviews, but a direct examination of coefficients for a few lemmas such as *bargain*, *buy*, *cheap*, *expensive*, and *frivolous* suggests otherwise. In fact, *bargain*, *buy*, and *cheap* are all marginally associated with male labels (0.091, 0.099, and 0.092), and *expensive* and *frivolous* are both marginally associated with female labels (0.082, 0.11). Meanwhile, the lemma *pleasure*, which Radway frames as a proxy for a certain kind of reading, has a slightly stronger female-label-informing coefficient (0.17). Radway’s argument, that some objected to or feared the ease with which certain texts or modes of production (like the Harvard Classics) transgressed social and rhetorical boundaries separating culture and commerce, may be relevant here. Perhaps book reviews were more likely to cross these boundaries without making direct reference to commercial language. Rubin has argued convincingly that the book review was an implicitly commercial genre by the turn of the century, due to its close ties to publishing and bookselling. Indications of gendered differences in the subject matter of reviewed books suggest that book reviews published in *The New York Times* between 1905 and 1925 followed the news approach, and in this sense my results echo one of Rubin’s points. However, even the summary-heavy reviews in *The New York Times* show evidence that, behind a veneer of neutrality, gender norms were being established, parameterized, and linked to taste-making.

This paper's secondary objective, as described in my introduction, was to demonstrate that large scale analysis of book reviews has the potential to go further and do more. The cultural norming I have discussed is an effective example of the analytical work that might be undertaken, I believe, because there are clear compatibilities between prior work and the kinds of patterns that large scale analysis of book reviews is likely to reveal. In turn, existing scholarship on cultural mediation has evocative connections to broader questions of authorship, readership, and symbolic capital. The full range of cultural analytics research questions that large scale analysis of book reviews might supplement or reshape is, for now, a matter of speculation. Better developed metadata could enable questions about how review language norms differed based on the review's author, the genre of the reviewed work, book prices, or book publishers. With a still wider corpus of more periodicals, these analytical methods have the potential to address Collier's call for a better understanding of the object of knowledge at the center of periodical studies, and to make smaller scale studies speak more directly to one another. These inquiries could be widened to compare book reviews to other types of reader response and literary criticism, or to trace patterns between reviews and the stylistic features of the books they respond to. As is the case with so many topics in cultural analytics, possibilities abound.

NOTES

¹ Lucien Febvre, and Henri-Jean Martin, *The Coming of the Book: The Impact of Printing 1450-1800*, trans. David Gerard (New Left Books, 1976), 235.

² Frank Donoghue, *The Fame Machine: Book Reviewing and Eighteenth-Century Literary Careers* (Stanford University Press, 1996), 1.

³ Michael Gavin, *The Invention of English Criticism: 1650–1760* (Cambridge University Press, 2015), 20.

⁴ Joan Shelley Rubin, *The Making of Middlebrow Culture* (University of North Carolina Press, 1992), 17, 19.

⁵ Janice A. Radway, *A Feeling for Books: The Book-of-the-Month Club, Literary Taste, and Middle-Class Desire* (University of North Carolina Press, 1997), 145.

⁶ Book reviews, of course, continued to evolve after 1900. See Rubin for a description of major 20th-century trends. For an overview of book reviews in early American periodicals, see Frank Luther Mott, *A History of American Magazines: 1741-1850* (Harvard University Press, 1938), 54-56.

⁷ Patrick Collier, "What Is Modern Periodical Studies?," *The Journal of Modern Periodical Studies* 6.

2 (2016): 92–3, <https://doi.org/10.5325/jmodeperistud.6.2.0092>.

⁸ Andrew Piper, “There Will Be Numbers,” *The Journal of Cultural Analytics*, May 23, 2016. DOI: 10.22148/16.006.

⁹ Richard Kluger, qtd. in Rubin, 92.

¹⁰ I wish to extend Allison Hegel a note of gratitude, as her work on book review trends on Goodreads has been crucial to the development of my thinking. I would like to thank Richard Jean So for reading and responding to this essay in a draft stage. My wife Rebecca Lee also read several versions of this essay and offered numerous suggestions, for which I am deeply grateful. This work would not have been possible without the cooperation of *The New York Times*, and the dedicated staff at the University of Pittsburgh who helped facilitate the collaboration. Finally, I want to thank Andrew Piper for all his advice and encouragement developing the dataset for this article and revising it for publication.

¹¹ Ted Underwood, David Bamman, and Sabrina Lee, “The Transformation of Gender in English-Language Fiction,” *The Journal of Cultural Analytics* February 13, 2018. DOI: 10.7910/DVN/TEGMGI.

¹² Ibid.

¹³ Here I am invoking Gérard Genette’s idea of a *paratext*, as well as his distinctions between *authorial* and *allographic*. See Genette, *Paratexts: Thresholds of Interpretation* (Cambridge: Cambridge University Press, 1997), 2-12.

¹⁴ The engaged and sustained discussion of these subjects in and around periodical studies foregrounds much of my work on *The New York Times Book Review*, including Dallas Liddle, “Genre: ‘Distant Reading’ and the Goals of Periodicals Research.” *Victorian Periodicals Review* 48.3 (October 2, 2015): 383–402.

<https://doi.org/10.1353/vpr.2015.0047>; Hannah McGregor, “Remediation as Reading: Digitising *The Western Home Monthly*.” *Archives and Manuscripts* 42.3 (September 2, 2014): 248–57.

<https://doi.org/10.1080/01576895.2014.958864>; and Jeffrey Drouin, “Close- And Distant-Reading Modernism: Network Analysis, Text Mining, and Teaching *The Little Review*.” *The Journal of Modern Periodical Studies* 5.1 (August 13, 2014): 110–35. I wish to express my thanks to the reviewer who called my attention to several of the periodicals studies articles cited here and elsewhere in the article.

¹⁵ This growing area of overlap also includes (but is not limited to) Katherine Bode, *Reading by Numbers: Recalibrating the Literary Field* (Anthem Press, 2014); Katherine Bode, *A World of Fiction: Digital Collections and the Future of Literary History* (University of Michigan Press, 2018); Ryan Cordell, “Reprinting, Circulation, and the Network Author in Antebellum Newspapers,” *American Literary History* 27.3 (September 1, 2015): 417–45, <https://doi.org/10.1093/alh/ajv028>; Andrew Piper and Richard Jean So, “Women Write About Family, Men Write About War,” *The New Republic*, April 8, 2016. <https://newrepublic.com/article/132531/women-write-family-men-write-war>.

¹⁶ I have not positioned my work in relation to computational approaches to book reviews for, as Peter Boot argues, “Computational linguistic research as a tool for understanding the variety of online book response is still at a very early stage of development” (34). For an extended literature review of what had been published through 2013, see Boot, “The Desirability of a Corpus of Online Book Responses.” *Proceedings of the Workshop on Computational Linguistics for Literature*. ACL, 2013.

¹⁷ Pierre Bourdieu, *Distinction: A Social Critique of the Judgement of Taste*, trans. Richard Nice (Routledge, 2013), 223. Pierre Bourdieu, *The Field of Cultural Production: Essays on Art and Literature*, trans. Randal Johnson (New York: Columbia University Press, 1993), 30.

¹⁸ See, especially, Beth Driscoll, *The New Literary Middlebrow: Tastemakers and Reading in the Twenty-First Century* (Springer, 2014), 12–17.

¹⁹ Cecilia Konchar Farr and Tom Perrin, “Introduction: Inventing the Middlebrow,” *Post45*, The Middlebrow: A Special Cluster. July 2, 2016. <http://post45.research.yale.edu/2016/07/introduction-inventing-the-middlebrow/>. On the origins of the terms *highbrow* and *lowbrow* and tie to racist discourse, see Lawrence W. Levine, *Highbrow/Lowbrow: The Emergence of Cultural Hierarchy in America*. (Cambridge: Harvard University Press, 1990): 221–222. For additional documentation on the first uses of the term *middlebrow*, see Driscoll and Radway.

²⁰ Radway, 152.

²¹ Radway, 144.

²² Radway, 152.

²³ See, especially, Nina Baym, *Novels, Readers, and Reviewers: Responses to Fiction in Antebellum America*. (Cornell University Press, 1984).

²⁴ Rubin, xviii.

²⁵ Rubin, 35.

²⁶ Rubin, 37.

²⁷ Jaime Harker, *America the Middlebrow: Women's Novels, Progressivism, and Middlebrow Authorship between the Wars* (University of Massachusetts Press, 2007), 16.

²⁸ Driscoll, 17.

²⁹ Driscoll, 29.

³⁰ Radway, 144.

³¹ Maria DiCenzo, "Remediating the Past: Doing 'Periodical Studies' in the Digital Era." *ESC: English Studies in Canada* 41.1 (September 24, 2015): 19–39. <https://doi.org/10.1353/esc.2015.0000>.

³² See "Books and Authors," *The New York Times*, November 11, 1896, 6. <https://timesmachine.nytimes.com/timesmachine/1896/11/01/106850832.html>; "New Publications," *The New York Times*, May 17, 1896, 27. <https://timesmachine.nytimes.com/timesmachine/1896/05/17/105750114.html>; and "Bourget's New Novel Figures in Caillaux Murder Trial," *The New York Times*, July 26, 1914, 2. <https://timesmachine.nytimes.com/timesmachine/1914/07/26/100099730.html>

³³ Francis E. Brown, *The Story of The New York Times Book Review*. (The New York Times Company, 1968): 7, 9, 18.

³⁴ See, for example, *The New York Times*, July 15, 1905, 18. <https://timesmachine.nytimes.com/timesmachine/1905/07/15/101828118.html?pageNumber=18>

³⁵ *The New York Times*, January 22, 1922. <https://timesmachine.nytimes.com/timesmachine/1922/01/22/98973696.html>

³⁶ The dataset I have assembled divides several of these clusters into individual reviews in order to boost representation of reviews by perceived-female authors. Code and lemma frequency tables for this article are online at <https://github.com/mjlavin80/nyt-reviews-1905-1925-gender>.

³⁷ A work with multiple authors, all perceived to belong to one gender, were categorized as male-labeled-reviews or female-labeled-reviews respectively.

³⁸ Johanna Drucker, "Why Distant Reading Isn't," *PMLA* 132.3 (May 2017), 631.

³⁹ *The New York Times* Article Search API includes links to pdf files that can be downloaded. No "recognized text" is stored within these files, so I worked with staff at *The New York Times* license OCR data.

⁴⁰ I used Spacy (Python) for tokenization and lemmatization. See Michal Toman, Roman Tesar, and Karel Jezek, "Influence of Word Normalization on Text Classification," *Proceedings of InSciT* 4 (2006): 354–58.

⁴¹ I used PyEnchant to drop all non-dictionary matching terms from the corpus rather than replace them with their likeliest alternatives.

⁴² For a description of TF-IDF and its effects, see David M. Blei, Andrew Y. Ng, Michael I. Jordan, "Latent Dirichlet Allocation" *Journal of Machine Learning Research* 3 (January 2003): 993–1022.

⁴³ Amanda Golden and Cassandra Laity, "Feminist Modernist Digital Humanities," *Feminist Modernist Studies* 1 (September 2, 2018): 205, <https://doi.org/10.1080/24692921.2018.1503786>.

⁴⁴ *Ibid.*

⁴⁵ Catherine D'Ignazio and Lauren Klein, "Introduction," *Data Feminism*. Public comment draft (MIT Press, 2018) n.p. <https://bookbook.pubpub.org/pub/dgv1622>.

⁴⁶ See Matthew L. Jockers and Gabi Kirilloff, “Understanding Gender and Character Agency in the 19th Century Novel,” *The Journal of Cultural Analytics* Dec. 1, 2016. DOI: 10.31235/osf.io/sw85m; Eve Kraicer and Andrew Piper, “[Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction](#),” *Journal of Cultural Analytics*. January 30, 2018. DOI: 10.31235/osf.io/4kwrg; Susan Brown and Laura Mandell, “The Identity Issue,” *Journal of Cultural Analytics*. Feb. 13, 2018. DOI: [10.31235/osf.io/9qvby](https://doi.org/10.31235/osf.io/9qvby); Richard Jean So, Hoyt Long, and Yuancheng Zhu, “Race, Writing, and Computation: Racial Difference and the US Novel, 1880-2000.” *The Journal of Cultural Analytics* January 11, 2019. DOI: [10.22148/16.031](https://doi.org/10.22148/16.031); and Underwood, Bamman, and Lee.

⁴⁷ See Brendan O’Connor, David Bamman, and Noah Smith, “Computational Text Analysis for Social Science: Model Assumptions and Complexity,” *Proceedings of the NIPS Workshop on Computational Social Science and the Wisdom of Crowds*, December 1, 2011, <http://repository.cmu.edu/lti/212>; Dan Jurafsky et. al., “Linguistic Markers of Status in Food Culture: Bourdieu’s Distinction in a Menu Corpus,” *The Journal of Cultural Analytics* October 18, 2016. DOI: 10.22148/16.007; and Underwood, Bamman, and Lee, 2018.

⁴⁸ Female-labeled-reviews were oversampled, in part, to make the machine learning model less vulnerable to influence by the reviews of any particular year. When constructing my training and test set, I also used random samples of reviews from 1905 and 1925 to ensure that material from these two years did not overwhelm the predictive models.

⁴⁹ For an explanation of F1 scores and why they are a preferred measure, see Kevin P. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press, 2012).

⁵⁰ Likewise, because of multicollinearity, we cannot assume a feature with a high coefficient score, if used in isolation, will have robust predictive power. For more information on complexities of multicollinearity, see Yasin Asar, “Some New Methods to Solve Multicollinearity in Logistic Regression,” *Communications in Statistics - Simulation and Computation* 46.4 (April 21, 2017): 2576–86, <https://doi.org/10.1080/03610918.2015.1053925>. See also David W. Hosmer Jr, Stanley Lemeshow, and Rodney X. Sturdivant, *Applied Logistic Regression* (John Wiley & Sons, 2013).

⁵¹ Class weighting does not have any effect on how the list of term coefficients is assembled. The most informative coefficients are trained in a separate phase of the model. See “class weight” at https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

⁵² Barbara Hochman, *Getting at the Author: Reimagining Books and Reading in the Age of American Realism* (University of Massachusetts Press, 2009), 9-10.

⁵³ Simone de Beauvoir, *The Second Sex*, trans. Constance Borde (Vintage, 2011), 5.

⁵⁴ Moshe Koppel, Shlomo Argamon, and Anat Rachel Shimoni, “Automatically Categorizing Written Texts by Author Gender,” *Literary and Linguistic Computing* 17.4 (November 1, 2002): 401–12, <https://doi.org/10.1093/lrc/17.4.401>.

⁵⁵ Out of 129 signed reviews of work by women in the inventory I conducted, 97 were written by men and 32 by women, and those 32 reviews were written by only 22 different female reviewers. This gloss of review authorship is far from definitive, and reviewer gender may have had a greater effect in different periodicals, or at different times. More work on the extent of entanglement between the gender signal of a book reviewer and the gender signal of the reviewed author would be valuable.

⁵⁶ Randal Johnson, “Editor’s Introduction” to Pierre Bourdieu, *The Field of Cultural Production: Essays on Art and Literature*, trans. Randal Johnson (New York: Columbia University Press, 1993), 5.

⁵⁷ Pierre Bourdieu, *The Logic of Practice*, trans. Richard Nice (Stanford University Press, 1990), 73.

⁵⁸ Bourdieu, *Logic*, 72.

⁵⁹ Bourdieu, Field, 30. See also Kate Huppatz, *Gender Capital at Work: Intersections of Femininity, Masculinity, Class and Occupation* (Palgrave Macmillan, 2012).

⁶⁰ Relevant schools of thought include lexical field theory, generative grammar, and relational semantics. See Dirk Geeraerts, *Theories of Lexical Semantics* (Oxford University Press, 2010), which summarizes many of the touchstones in these subject areas.

⁶¹ Benjamin M. Schmidt, “Words Alone: Dismantling Topic Models in the Humanities” *Journal of Digital Humanities* 2.1 (2012). Accessed March 22, 2018, <http://journalofdigitalhumanities.org>.

⁶² Ibid.

⁶³ According to Wordnet’s documentation, “Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations.” For more information, see “What is Wordnet,” <https://wordnet.princeton.edu/>

⁶⁴ The noun *childs* may look like a tagging error, but it is in fact a proper noun, referring to “a waitress in a Childs restaurant.” See “America’s Short Story Writers,” *The New York Times*, February 6, 1921, BR2, BR21. <https://timesmachine.nytimes.com/timesmachine/1921/02/06/112670148.html?pageNumber=26>

⁶⁵ Descriptive statistics for the top 50 coefficients for male and female labels can be found in the Github repository for this article, in the folder “tables-and-figures,” in the files called ‘top_50_coef_mf_lemma_word_pos_data.csv’, ‘top_50_coef_mf_doc_count_data.csv’, and ‘senses.md’.

⁶⁶ “Notable Books in Brief Review,” *The New York Times*, August 12, 1917, BR2.

<https://timesmachine.nytimes.com/timesmachine/1917/08/12/102359453.html?pageNumber=20>

⁶⁷ “One of America’s Noble Women: Julia Ward Howe’s Biography by Her Daughters Reveals Stimulating Influence of a Fine Personality,” *The New York Times*, March 19, 1916, 101.

<https://timesmachine.nytimes.com/timesmachine/1916/03/19/99439021.html?pageNumber=38>

⁶⁸ Joseph, O’Connor, “Lord Dufferin’s Career: Sir Alfred Lyall’s Authorized Biography of One of the Most Famous British Statesmen and Diplomats of Recent Times,” *The New York Times*, April 1, 1905, BR1, BR2.

<https://timesmachine.nytimes.com/timesmachine/1905/04/01/101322196.html?pageNumber=21>

⁶⁹ Underwood, Bamman, and Lee point to Nancy Armstrong’s. *Desire and Domestic Fiction: A Political History of the Novel* (Oxford University Press, 1990) and Gaye Tuchman and Nina E. Fortin, *Edging Women Out: Victorian Novelists, Publishers and Social Change* (Routledge, 2012). See also Monika Maria Elbert, *Separate Spheres No More: Gender Convergence in American Literature, 1830-1930* (University of Alabama Press, 2000); T. Foster, *Transformations of Domesticity in Modern Women’s Writing: Homelessness at Home* (Springer, 2002); Kristin J. Jacobson, *Neodomestic American Fiction* (Ohio State University Press, 2010); Sally Ledger, *The New Woman: Fiction and Feminism at the Fin de Siècle* (Manchester University Press, 1997); and Francesca Sawaya, *Modern Women, Modern Work: Domesticity, Professionalism, and American Writing, 1890-1950* (University of Pennsylvania Press, 2011).

⁷⁰ Jockers and Kirilloff, 2016.

⁷¹ Underwood, Bamman, and Lee, 2018.

⁷² Ibid.

⁷³ Piper and So, <https://newrepublic.com/article/132531/women-write-family-men-write-war>

⁷⁴ In all three of these figures, an adjustment of +/- 0.05 is applied to counts to make it easier to differentiate male and female labels.

⁷⁵ Driscoll, 17.