

The Transformation of Gender in English-Language Fiction

Ted Underwood, David Bamman, and Sabrina Lee¹

Preprint of an essay to appear in *Cultural Analytics*, February 2018

This essay explores the changing significance of gender in fiction, asking especially whether its prominence in characterization has varied from the end of the eighteenth century to the beginning of the twenty-first. We have reached two conclusions, which may seem in tension with each other. The first is that gender divisions between characters have become less sharply marked over the last 170 years. In the middle of the nineteenth century, very different language is used to describe fictional men and women. But that difference weakens steadily as we move forward to the present; the actions and attributes of characters are less clearly sorted into gender categories. On the other hand, we haven't found the same progressive story in the history of authorship. In fact, there is an eye-opening, under-discussed 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. The number of characters who are women or girls also drops. We are confronted with a paradoxical pattern. While gender roles were becoming more flexible, the space actually allotted to (real, and fictional) women on the shelves of libraries was contracting sharply. We explore the evidence for this paradox and suggest a few explanations.

This essay considers both the gender positions ascribed to authors as biographical personages, and the signs of gender they used in producing characters. In both cases, we understand gender as a conventional role that people were expected to assume in order to become legible in a social context. Authors and characters have

¹ The evidence used in this paper has depended heavily on the labor of other hands. The HathiTrust corpus we use was processed by Boris Capitanu. The Chicago Novel Corpus was collected by Hoyt Long and Richard Jean So, and enriched with metadata by Teddy Roland. Conversation with Heather Love, Laura Mandell, and Allen Riddell, and the whole NovelTM research group, turned up valuable leads. Funding for this project was provided by SSHRC via NovelTM, directed by Andrew Piper, and by the Andrew W. Mellon Foundation via the WCSA+DC project, directed by Stephen J. Downie.

been coded according to a tripartite scheme (feminine / masculine / other or unknown), because that scheme organized most public representation of gender in the period we are studying.² Gender can certainly be more complex than these categories suggest, and flexible ontologies can be designed to illuminate the complexity. But this essay is inquiring about the history of conventional roles, not about the truth of personal identity, or the underlying processes that produce gendered behavior. We will cast light on those questions only indirectly, by showing that public signs of gender had a fluctuating and uncertain significance.

To trace the representation of character across 104,000 books, we needed a way to identify the characters in a work and separate them from each other. We used a pipeline called BookNLP,³ which identifies character names in a work of fiction and clusters those names, so that “Elizabeth” and “Elizabeth Bennet” are linked as a single person. Then it identifies words that are connected to each character in a range of different ways: because they're actions she performs, actions of which she's the object, adjectives that modify her, or nouns that she governs (“her spirits,” for instance).

Some aspects of this pipeline are relatively fragile. In using only proper names to identify characters, we miss out on characters who are exclusively referred to in generic terms, such as “the footman”; uncommon nicknames (such as “Pip” for Philip Pirrip) may result in a single character sometimes being divided into multiple roles. Moreover, a role like “Miss Bennet” can belong to different people over the course of a narrative. Gender, on the other hand, is comparatively easy to recognize. Even when a single character is split into multiple roles, BookNLP usually assigns gender to each role correctly, guided by names and honorifics (“Mr,” “Lady,” “Miss”). First-person narrators constitute a significant exception, because there are no signs of gender attached to the pronoun “I.” So first-person narrators won't be considered in this study; further research will be needed to determine if they have a different history.

² The words “woman” and “man” are ambiguous; they can be taken as referring to public social roles, private identities, or biological genotypes. In this essay, we are always referring to public roles defined by a particular social context.

³ David Bamman, Ted Underwood, and Noah Smith, “A Bayesian Mixed Effects Model of Literary Character,” *ACL 2014*, <http://www.cs.cmu.edu/~ark/literaryCharacter/>.

Where other characters are concerned, BookNLP assigns gender with reasonable accuracy. Women are identified with 94.7% precision/83.1% recall and men are identified with 91.3% precision/85.7% recall. More importantly for the diachronic argument we will be making in this essay, accuracy seems to be relatively stable over time.

In short, while BookNLP may not provide a precise count of the sheer number of characters in a volume, it does consistently divide descriptions of men from descriptions of women. This makes it possible to pose several interesting questions. To start with a simple and obvious one: How much space did writers devote to men and women, respectively?

The masculinization of fiction, 1800-1960

We are working with a collection of 104,000 works of fiction. They are spread over 306 years, from 1703 to 2009, but the vast majority are dated from 1780 to 2007, and that's where we will focus our attention. The works we use are drawn mostly from HathiTrust Digital Library, although we have also compared that collection to an alternate one developed at the Chicago Text Lab, and covering the period 1880-1989.⁴ We can subset these collections in various ways, to emphasize authors who are relatively prominent or obscure, popular or elite. But broadly, and taken as a whole, our evidence is shaped by the book-buying practices of academic libraries (with additional contributions from the Library of Congress and New York Public Library). It doesn't include everything published in the period, and certain important sites of publication (pulp magazines, for instance) are known to be underrepresented. On the other hand, we have checked this collection against a less academic sample drawn from *Publishers Weekly*, and (as we will explain shortly) the most dramatic trends affecting authorship seem to be broadly the same in both samples.

In any case, a collection drawn from academic libraries can certainly tell us a

⁴ To make our argument reproducible, we have shared data and code at <https://github.com/tedunderwood/character>.

lot about the literary tradition that has been studied in universities. Going on academic accounts of literary history, we might predict that the prominence of female characters in this tradition would have increased in the nineteenth and twentieth centuries—with, to be sure, some well-known interruptions. There is a notorious backlash against first-wave feminism in the middle of the twentieth century, for instance—dated by Gayle Greene to the 1940s and 50s.⁵ But we might hope at least to see an overall story of progress across two centuries.

If we start with that loosely positive expectation, we will be disappointed. In fact, from the nineteenth century through the early 1960s we see a story of steady decline. In Figure 1, we have plotted the proportion of words used in characterization that describe women. “Words used in characterization” here includes verbs a character governed, and nouns they possessed (like “spirits”), as well as adjectives that modify a character. Dialogue spoken by characters is not counted, but including it would not materially change the pattern. Characters of unknown gender have been excluded from the total, so the proportion being plotted (here, and elsewhere in the article) is simply the ratio of words that describe women to words that describe either men or women. This proportion actually declines from the middle of the nineteenth century to the middle of the twentieth — in the very period when we might expect to see the effects of first-wave feminism. The trend reverses around 1970, for reasons we will investigate later.

⁵ Gayle Greene, *Changing the Story: Feminist Fiction and the Tradition* (Bloomington: Indiana University Press, 1991), 39-41.

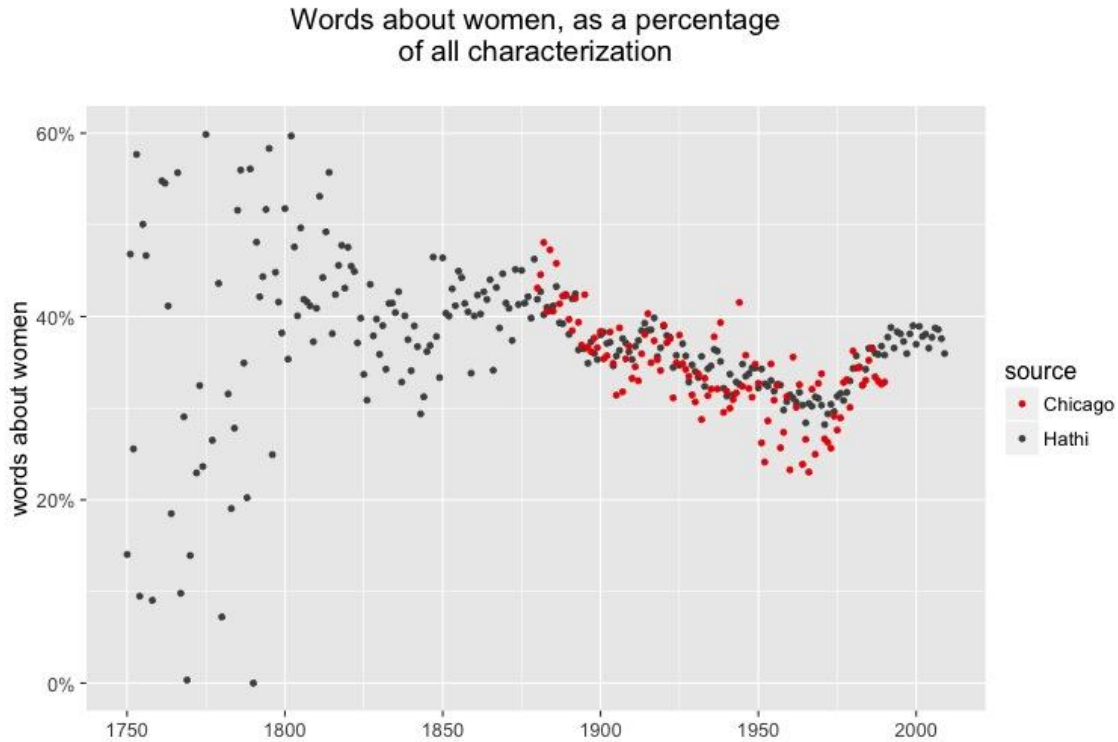


Figure 1. *The percentage of words used in characterization that describe women.*

This picture is counter-intuitive enough that our first instinct was to ask whether it might be an artefact of some error in our collection or methods. We have already acknowledged that this inquiry is mostly limited to university libraries, but could there also be some other distortion in the way we have chosen volumes from those libraries? One check is provided by the fact that we have selected volumes in two very different ways, from two different sources. The HathiTrust fiction corpus was separated from nonfiction algorithmically and includes one copy of every volume that the algorithm tagged as fiction (including works in translation and folktales, for instance). The Chicago Novel Corpus was selected manually and includes only novels composed in English, emphasizing American works identified as prominent (because they were held by *many* academic libraries). Although we attempted a rough deduplication of HathiTrust, the Chicago corpus is also less likely to include duplicates and is far better at dating works by their date of first publication. But these differently-constructed corpora display broadly the same pattern of gender inequality

in characterization.

In short, this is a real trend. But how can we explain a trend that runs directly against our assumptions about social progress? One answer is that, during the period when women were becoming less prominent as characters in fiction, women *writers* were also losing shelf space in the library. As we will see in a moment, women invent female characters much more often than men do, so any decline in the number of women writers will create a corresponding decline in description of women. And there was, in fact, a fairly stunning decline in the proportion of fiction writers who were women, from the middle of the nineteenth century to the middle of the twentieth.

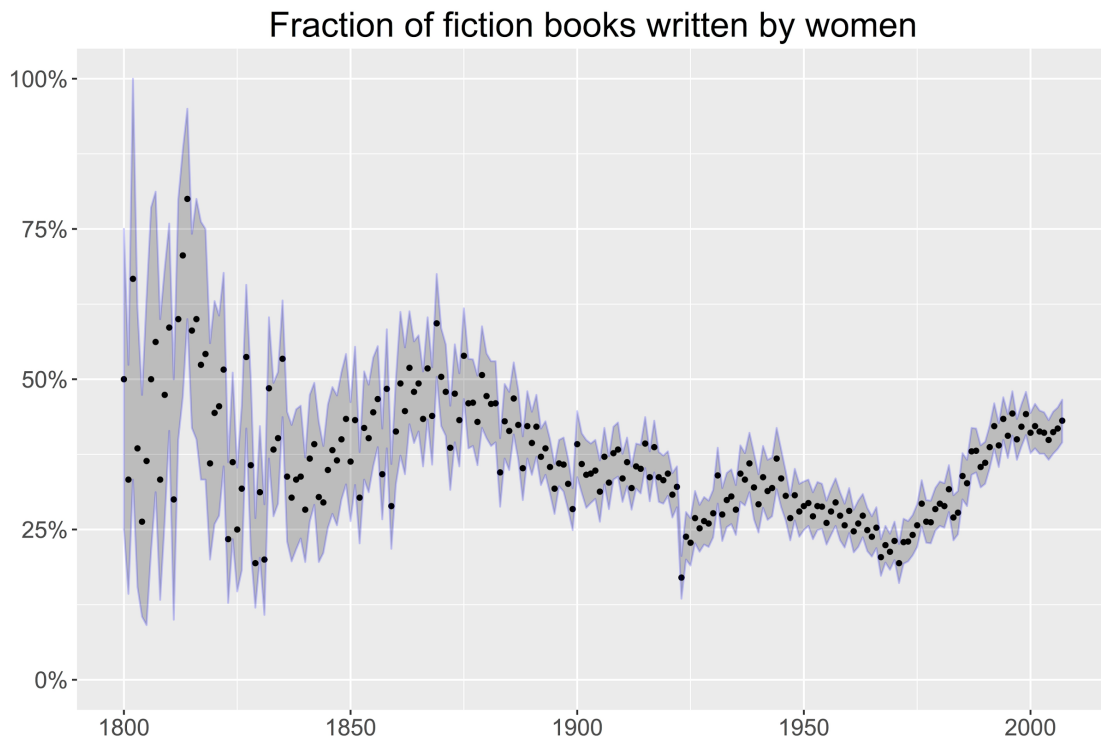


Figure 2. The fraction of English-language volumes of fiction in HathiTrust written by women. Dots are the fraction actually recorded for a given year of publication; the shaded area represents a 95% confidence band calculated by bootstrap resampling.

Women go from representing almost half the authors of fiction, to barely a quarter. If this trend is real, it is an important fact about literary history that ought to be foregrounded even, say, in anthology introductions. But the story has not been

widely publicized. There are some existing works of scholarship that highlight pieces of the decline. The most important of these is *Edging Women Out*, where Gaye Tuchman and Nina Fortin report that “before 1840 at least half of all novelists were women; by 1917 most high-culture novelists were men.”⁶ Separately, scholars of modernism have traced a redefinition of high culture that tended to disadvantage women. Suzanne Clark has explained that “modernism reversed the increasing influence of women’s writing” by defining itself against the sentimental tradition.⁷ But it is telling that Clark interprets this shift as a reversal of nineteenth-century progress for women. It appears that scholars of each period are able to see the possibility that female authorship was declining in their own period. But no one has been willing to advance the dismal suggestion that the whole story from 1800 to 1960 was a story of decline—at least not until Chris Forster noticed some clues in HathiTrust two years ago.⁸ Franco Moretti suggested that the late-nineteenth-century decline might just be one dip in a regular generational undulation.⁹ Nor has everyone found the evidence for decline persuasive even in the late nineteenth century. Ellen Miller Casey examined late-nineteenth-century reviews in *The Athenaeum* and concluded that “there is little evidence of a steady male invasion edging women out.”¹⁰

So how trustworthy is our own evidence? There are two reasons for skepticism. One is that the novels counted in Figure 2 are drawn mostly from academic libraries. How well do those collections represent the wider world of fiction? Book historians are often doubtful.¹¹ (Tuchman and Fortin may have restricted their claim to “*high-culture* novelists” because they anticipated a similar skepticism.) A second problem is that to assign authorial gender to 104,000 volumes we had to rely on algorithmic inference,

⁶ Gaye Tuchman and Nina E. Fortin, *Edging Women Out: Victorian Novelists, Publishers, and Social Change* (London: Routledge, 2012), 7

⁷ Suzanne Clark, *Sentimental Modernism: Women Writers and the Revolution of the Word* (Bloomington: Indiana University Press, 1991), 1.

⁸ Chris Forster, “A Walk Through the Metadata: Gender in the HathiTrust Dataset.” September 8, 2015. <http://cforster.com/2015/09/gender-in-hathitrust-dataset/>

⁹ Franco Moretti, *Graphs, Maps, Trees: Abstract Models for Literary History* (London: Verso, 2005), 27.

¹⁰ Ellen Miller Casey, “Edging Women Out? Reviews of Women Novelists in the *Athenaeum*,” *Victorian Studies* 39 (1996): 154.

¹¹ Katherine Bode, “The Equivalence of ‘Close’ and ‘Distant’ Reading; or, Toward a New Object for Data-Rich Literary History,” *Modern Language Quarterly*, 78.1 (2017): 77-106.

using names recorded in the United States Census as a guide.¹² What about ambiguous names, or non-European names, or pseudonyms, or multiple authors?

To address these doubts, we decided to construct a second sample, using independent sources and methods. From 1873 forward, *Publishers Weekly* recorded books published in America. This may not be an exhaustive list of publications, but it definitely includes many things not digitized or preserved in academic libraries. (In fact only 56% of the works of fiction we sampled from *Publishers Weekly* are contained in HathiTrust.) We manually sampled four years to see whether gender trends in this larger domain would echo trends in academic digital libraries. Since we were using our own judgment to assign authors to a gender category, this sample also addressed potential concerns about the accuracy of algorithmic inference based on names.

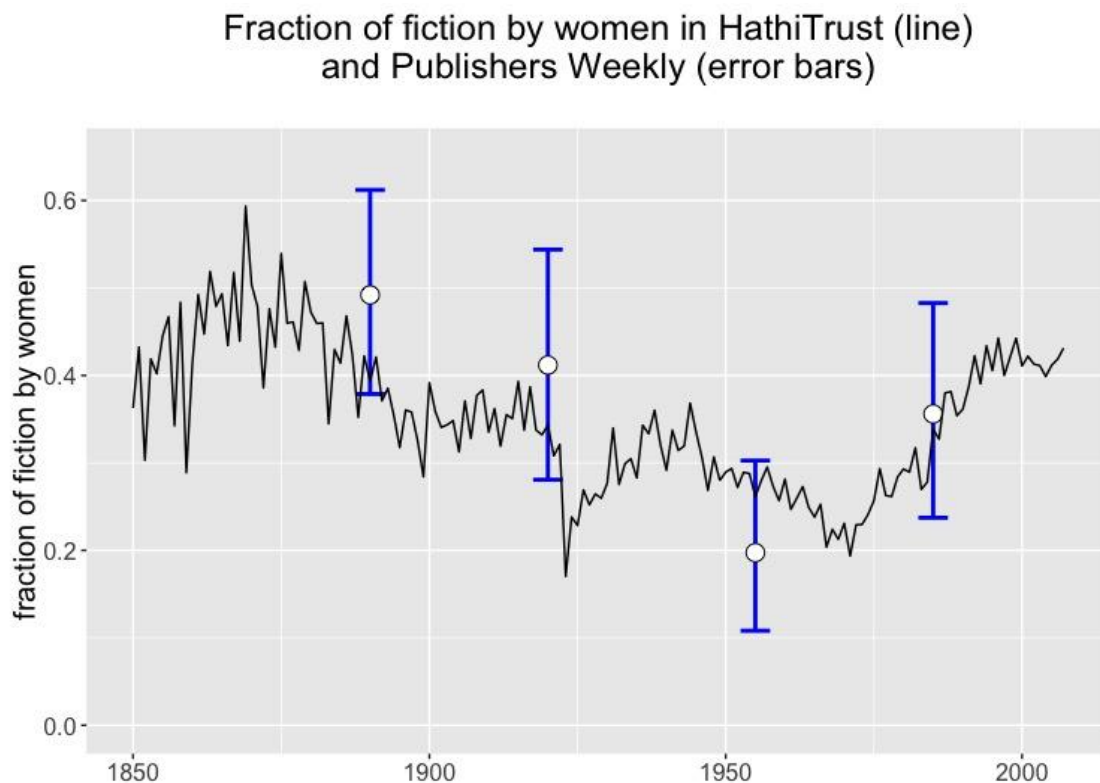


Figure 3. Gender ratios in the HathiTrust sample (as a yearly point estimate) and 95% confidence intervals for four years sampled in Publishers Weekly.

¹² Inference from personal names was done by Gender-ID.py, created by Bridget Baird and Cameron Blevins, 2014, <https://github.com/cblevins/Gender-ID-By-Time>.

Manual sampling is labor-intensive, so our samples are not huge, and our error bars leave considerable wiggle room. But it is safe to say that we have found no evidence that the broad trends in HathiTrust are produced merely by library purchasing patterns. Minor fluctuations, such as the suspiciously sharp drop at 1923, for instance, may well be library-specific. But the broad decline in the proportion of fiction written by women is, if anything, even more dramatic in this larger sample.

There is one final source of skepticism to address. Some critics of distant reading complain that library-based samples are too narrow. But others object that they are too broad. Do we really need to care about every novel ever published? Shouldn't we care more about novels with cultural significance? So we checked the yearly fiction bestsellers reported by *Publishers Weekly*. This sample covers a shorter timeline and displays slightly different patterns, but here too, the prominence of women in fiction broadly declines from the first half of the twentieth century to the period 1950-1970. (To tell the truth, checking bestsellers is probably overkill. A massive change in academic libraries would be important in itself. But readers who have any doubts about the significance of libraries can set their doubts aside: this was a broad trend, and it had some impact on every aspect of literary life.)

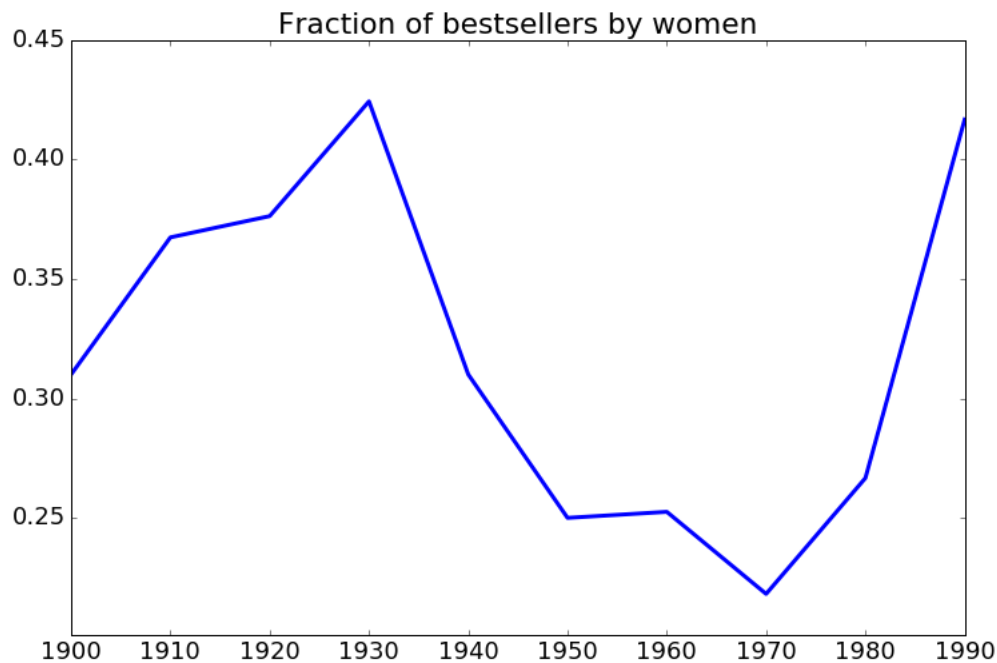


Figure 4. The authors of yearly bestsellers from *Publishers Weekly*.

So why did it happen? Tuchman and Fortin offer several hypotheses. First, they emphasize that early-nineteenth-century fiction was dominated by women because novel-writing was not yet a high-status career. “[A]fter 1840 some men may have become novelists because writing fiction increasingly brought status. Additionally, after 1840 the job conditions of novelists improved” (9). Gentrification stories of this sort are familiar. Talia Schaffer has pointed out that male aesthetes similarly moved in on descriptions of dress and interior décor that had been seen as feminine—until they became too prestigious to leave to women.¹³ Secondly, Tuchman and Fortin trace social pressures in literary reviewing, and in the terms of publishers’ contracts, that subjected women to increasing disadvantages in the later nineteenth century. (Ellen Miller Casey’s close analysis of *The Athenaeum* comes to an opposite conclusion, at least about reviews in one periodical [158-59].) Finally,

¹³ Talia Schaffer, *The Forgotten Female Aesthetes: Literary Culture in Late-Victorian England* (Charlottesville: University of Virginia Press, 2000), 23-29, 73-87.

Tuchman and Fortin also acknowledge that intellectual careers other than “novelist” were opening up to women. To fully map this expansion would require many different kinds of social evidence. But if we stick with the evidence of authorship, fiction is one of the few parts of the library where representation of women seems to have declined. If we consider all other categories, collectively, we see an enormous expansion (interrupted by the much briefer sort of pause in the middle of the twentieth century that we might have expected). It is not hard to see how expanding opportunities on this scale might have lured women away from the novel.

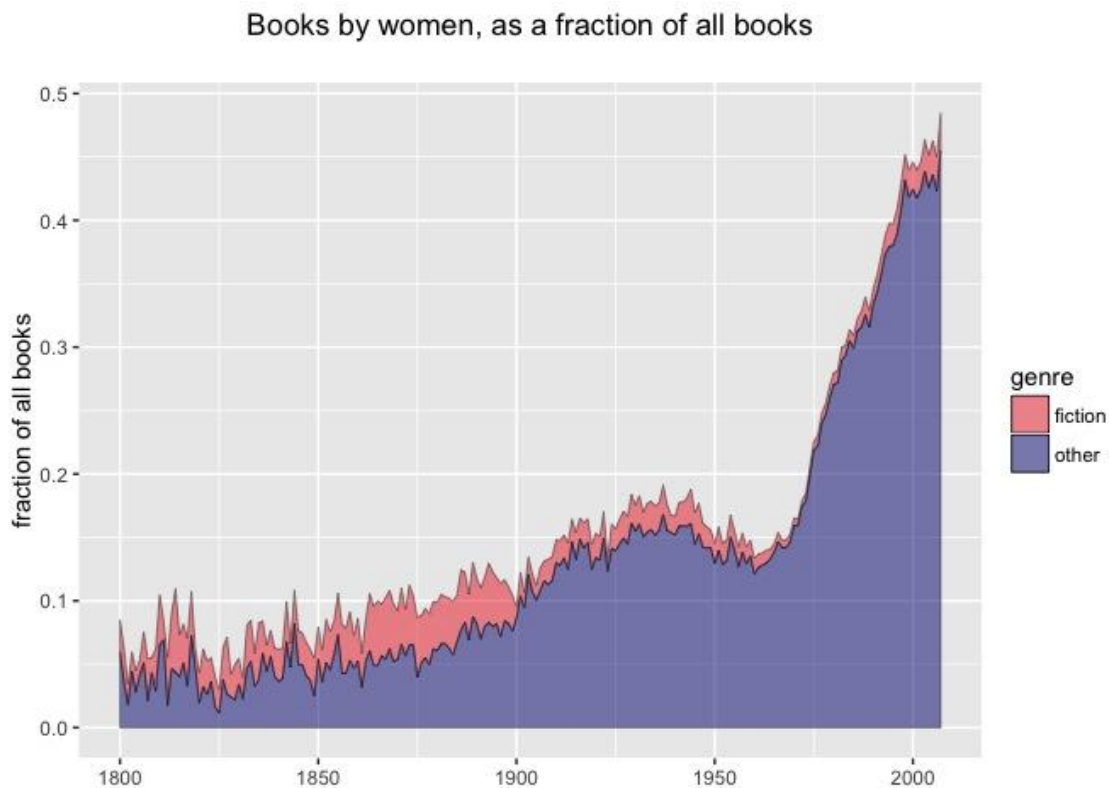


Figure 5. Books by women, as a fraction of all books in HathiTrust. The category “other” is mostly nonfiction, although it also includes small amounts of poetry and drama.

Social causality is never easy to untangle. Tuchman, Fortin, and Casey, for instance, come to opposite conclusions about the impact of nineteenth-century reviewing. In this paper, we won’t completely resolve the tension between competing explanations either. But the occupational shifts that Tuchman and Fortin have

described—male “gentrification” of the novel, and the opening of other careers for women—are very plausible explanations of the changes we see down to, say, 1940. After that point, a different pattern appears: participation in fiction and nonfiction move together, down and then (spectacularly) up, possibly in response to the broader fortunes of feminism. From 1940 forward, it appears that authorship of fiction is shaped less by social factors that guided women toward or away from fiction in particular than by broader attitudes toward femininity and work.

But this is not to imply that ideological pressures of that broad kind were absent before 1940. Evidence gathered from BookNLP does hint that the decline from 1850 to 1940 also had an imaginative or ideological component. The change wasn’t simply that more men decided to become novelists, or that women found other opportunities—but that fiction itself became more attentive to men. We can illustrate this by considering the space on the page that writers allot to fictional men or women. Women certainly write about women more than men do. In books written by men, women occupy on average only quarter to a third of the character-space.¹⁴ In books written by women, the division is much closer to equal. This gap between the genders is depressingly stable across two hundred years and may be, for some readers, the biggest story coming out of this inquiry.

¹⁴ The concept of character-space here is loosely adapted from Alex Woloch, *The One Versus the Many: Minor Characters and the Space of the Protagonist in the Novel* (Princeton: Princeton University Press, 2003).

Description of women, as a percentage of characterization,
broken out by author gender

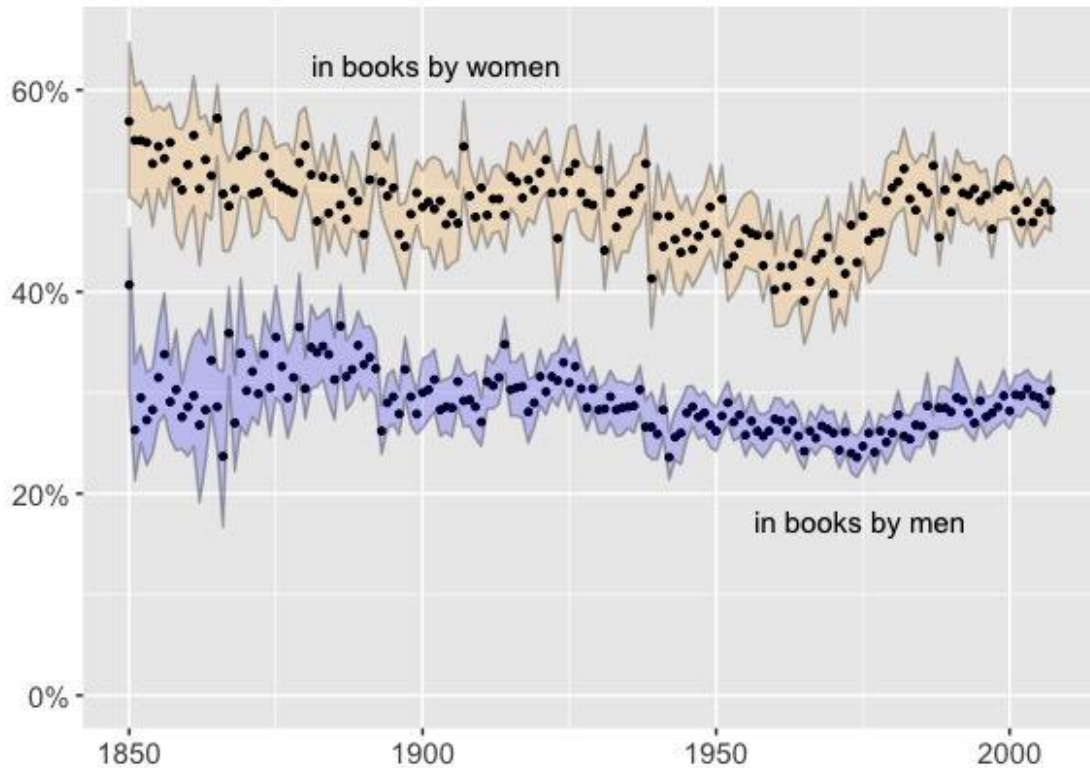


Figure 6. Words devoted to description of women, broken out by author gender. *HathiTrust* and the *Chicago Novel Corpus* are folded together here.

But there is also another, subtler story in this image, because the declining prominence of women as characters between 1850 and 1960 remains visible even after we separate volumes written by men and women. This suggests that the underrepresentation of fictive women inside books cannot be completely explained by the underrepresentation of women writers on the shelves of libraries. When we separate authors by gender, we find that women were becoming less prominent *even in books by women* across this century.

As we have said, this is only a subtle shift, and we are not yet sure how much should be inferred from it. Literary scholars often move rapidly to draw broad cultural conclusions from scattered traces: if we wanted to do that here, we could talk about a steady masculinization of fiction, stretching from the middle of the nineteenth century

to the apotheosis of Papa Hemingway in the 1950s. But in reality, more research is needed to understand what happened. For instance, how much of this change was due to shifting literary ideals (the masculinity of high modernism), and how much to the displacement of domestic fiction by adventure, detective, or Western genres? It is also important to remember that any “masculinization of fiction” between 1800 and 1960 took place against the backdrop of a broader social change that gave women a vastly larger proportion of the nonfiction book market than they had held in the nineteenth century. We are in the habit of reading literary change as a direct reflection of deep cultural shifts. But the relation between literature and other aspects of culture can also be indirect or even reversed: for instance, if women abandon the novel because they are finding wider opportunities elsewhere, fiction and nonfiction can move in opposite directions.

The instability of gender

Leaving the sheer allocation of space in books aside for now, let’s move to consider the characters themselves. Up to this point we have taken public signs of gender at face value, as if their meaning remained constant. But we know, in fact, that characters are gendered to different extents and in different ways. This is a complex topic, and we won’t be able to capture all its nuances on a scale of centuries. But we can ask, at least, how strongly a public, binary conception of gender shaped other aspects of characterization. Were fictive men quite different from fictive women, or were the differences between characters mostly unrelated to conventional signs of gender? And how did the answer to that question vary across the timeline?

Questions about the relative strength of different relationships are a place where quantitative methods shine. The only thing that makes this problem tricky is that we can’t define in advance which aspects of character might be surreptitiously gendered. Instead, we need a capacious way of representing lots of different aspects of character at once, in order to ask how well gender explains them collectively. A bag-of-words representation has worked well for many similar problems. We represent each character by the adjectives that modify them, the verbs they govern, and so on

— excluding words that are names of gendered roles like *girlhood* or *boyhood*, *wife* or *husband*, and personal names. Then we show a model some of the characters, labeled with grammatical gender. The model will learn what it means to be “masculine” or “feminine” purely by looking at the tacit assumptions expressed in words associated with the character. It can then use those patterns to make predictions about other characters it hasn’t yet seen. If the model turns out to be accurate, we can conclude that practices of characterization were powerfully organized by a binary conception of gender: even apparently innocuous words like (say) “smile” or “grin” appear to tacitly predict gender.¹⁵ If the model becomes less accurate, we can conclude that gender was becoming a less pervasive organizing structure—or at least, that gender was being expressed in ways that no longer aligned with the binary division between *he* and *she*.

When we attempt this comparison, a clear long-term pattern emerges: the implicit differences between masculine and feminine characters get steadily harder to discern from the middle of the nineteenth century to the beginning of the twenty-first.

¹⁵ For an analogous project reaching different conclusions, see Matthew Jockers and Gabi Kirilloff, “Understanding Gender and Character Agency in the 19th Century Novel,” *Cultural Analytics*, 2016, <http://culturalanalytics.org/2016/12/understanding-gender-and-character-agency-in-the-19th-century-novel/>. Jockers and Kirilloff find little change in the nineteenth century. They may not be wrong; most of the change we have detected comes in the twentieth century.

Accuracy of gender prediction, 1600-character samples

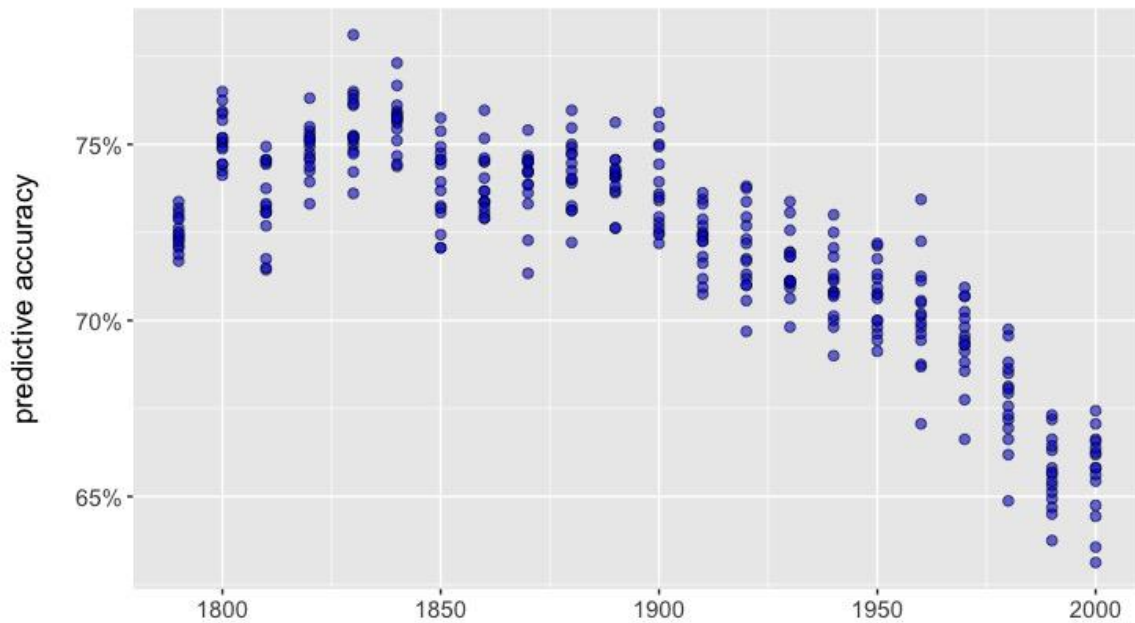


Figure 7. How easily can grammatical gender be predicted from ostensibly ungendered evidence? 84,000 characters drawn from novels in HathiTrust, sampled in groups of 800 men and 800 women.

The blue dots are scattered in columns because we ran fifteen different models in each decade, randomly selecting 1600 characters each time, and classifying them using the 2200 most common words in that group of characters.¹⁶ To make these comparisons apples-to-apples, we ensured that the median “size” of the text associated with characters was roughly 54 words in every decade. That’s not a lot of words (many characters are described fairly briefly), so these models are never as accurate as a model of genre (which uses a whole volume for each inference). Genre models often exceed 90% accuracy; these character models peak around 76%.

This picture may align somewhat better with expectations about progress than the first half of this article did. As we move from 1850 forward, characters are less and less clearly sorted along a masculine-feminine axis. That sounds like the kind of

¹⁶ The classification algorithm was regularized logistic regression. The feature count (2200) was arrived at by optimizing 1600-character models on the collection as a whole; hyperparameters were optimized separately for each model.

convergence we might have expected to see as the notion of separate spheres was challenged.¹⁷ Before 1850, patterns of change are less clear, both because we have fewer data points and because we have observed that the trend in this part of the timeline depends in practice on debatable assumptions—like the decision to exclude certain words (*girlhood*, *husband*) as tautological signs of gender. But if there were a slight increase in accuracy from 1780 to 1840, it would fit received narratives about domestic ideology reasonably well. Virginia Woolf, at any rate, believed that “the sexes drew further and further apart” as the nineteenth century began.¹⁸

Excluding certain gendered words as “tautological” clues is a debatable move, because it is not obvious where to stop. We may agree that *he* and *she* convey nothing but a gender signal. References to a character's *boyhood* are almost as clear. But *husband* and *wife* are more debatable cases. In this dataset, most of the characters with wives will be men, but that really depends on sexuality rather than gender. Probably the best solution would be to exclude gendered relationship terms as predicates (since “Mr Darcy was her husband” conveys Darcy's gender tautologically), but not as attributes possessed by the character (since possession of a husband signals Elizabeth's gender only if she happens to be heterosexual). In some modeling runs we made that distinction, but not in the final version represented above. Nor did we exclude references to the body or articles of dress, although skirts and mustaches are arguably signs of a (cis)gendered identity.

These fine lines are interesting philosophically, but the debatable cases make little difference to the biggest historical pattern in Figure 7: the decline in accuracy from 1840 to 2007. In the course of writing this article we have run many different tests on that pattern. We excluded, or didn't exclude, different sets of words. We used different sources of data (novels after 1923 were originally drawn from Chicago Text Lab, but the current version relies entirely on HathiTrust). We also tried different modeling strategies (feature sets and hyperparameters were optimized in different

¹⁷ Monika M. Elbert, ed., *Separate Spheres No More: Gender Convergence in American Literature* (Tuscaloosa: University of Alabama Press, 2000).

¹⁸ Virginia Woolf, *Orlando: A Biography*, ed. Rachel Bowlby (Oxford: Oxford University Press, 1992), 219.

ways; characters were selected differently). Finally, we tested the reliability of BookNLP's gender inference in different periods by comparing the predicted gender to manual judgments for 525 characters evenly distributed over time, to make sure the decline in accuracy wasn't merely a decline in the reliability of our ground truth (Figure 8). We have shared code and data to make our work reproducible, but the upshot is that we are very confident in the broad outlines of Figure 7. The blurring of gender boundaries from 1840 through the late twentieth century is quite robust; the trend is not likely to vanish with a different set of sampling or modeling choices.

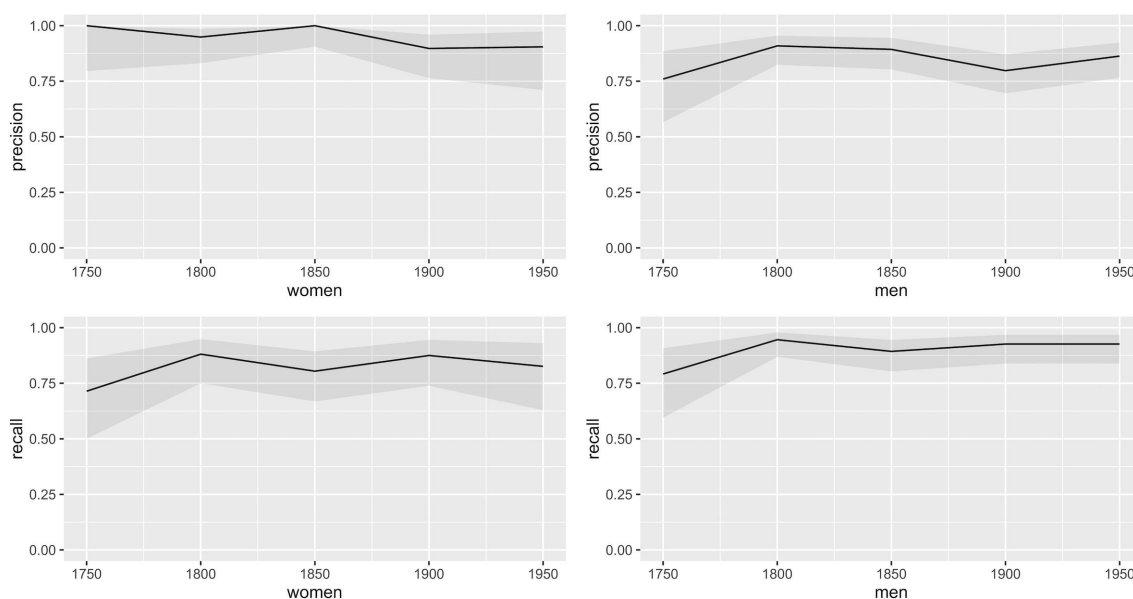


Figure 8. Accuracy of BookNLP's gender inference for characters. The only significant departure is in the eighteenth century—the wrong place to explain the twentieth-century decline in Figure 7.

Because the blurring of gender boundaries begins around the same time as the space allotted to description of women in fiction starts to shrink, it is tempting to assume that these two trends must have been related. But if so, the relationship is far from intuitively clear. Our models in each decade are balanced to include equal numbers of characters: 800 men and 800 women. So population-level changes in the ratio of fictive men to fictive women wouldn't directly affect the results.

What about changes in authorship? We have seen that women write a smaller proportion of works of fiction as we move into the twentieth century. That might

explain the growing blurriness of gender categories if, for instance, gender differentiation tended to be crisper in novels written by women. But the reverse is true (Figure 9). When we compare fictive men and women only *within* fiction written by women, or *within* fiction written by men, the models that predict the genders of characters created by women are consistently less accurate (by 2.5% on average). Gender differences seem to be drawn more starkly in stories written by men. It is not immediately clear why, although one might guess that the explanation is related to the underrepresentation of women in their imaginative worlds. The lone woman in a Western or detective story may tend to be depicted very strongly as “The Woman.” We might call this the Irene Adler Exception, if Katha Politt hadn’t already memorably dubbed it “the Smurfette Principle.”¹⁹

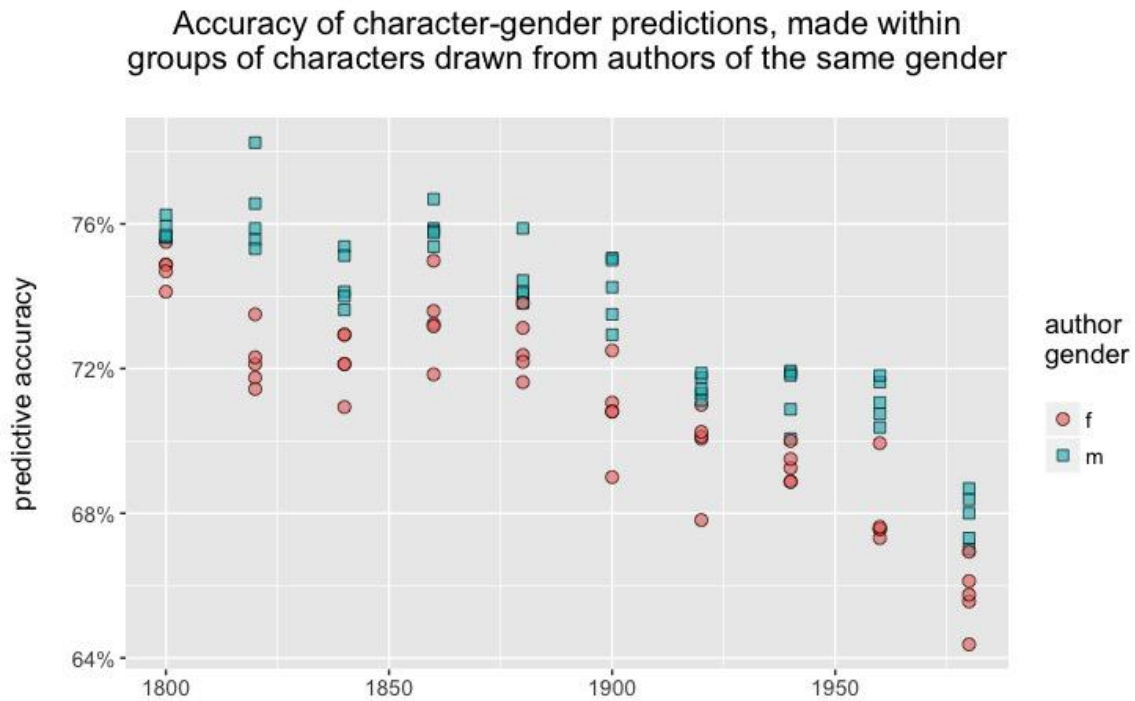


Figure 9. The accuracy of models that attempt to distinguish character gender in groups of characters drawn only from books by men, or by women. Classification parameters otherwise as in Figure 7.

¹⁹ Katha Politt, “Hers; The Smurfette Principle,” *New York Times*, April 7, 1991.

The accuracy of the models in Figure 7 varies by 10% or so across 230 years—from roughly 76% to 66%. Because these methods are new, it may be difficult to know intuitively whether that constitutes a dramatic change or a subtle one. And how strong is 76% accuracy to begin with? Is that a relatively low score, indicating that gender is never strongly marked? Or is it relatively high accuracy, given that the model has only 54 words of evidence for each character (on average), with some of the most obviously gendered words excluded?

Mathematics won't in itself give us an answer. To address this question, it may be useful to juxtapose another model trained on the same data. For instance, we could attempt to infer the gender of the *author* from the character description. To prevent author gender from acting as a mere proxy for the overrepresentation of male characters in books by men, we discarded some men from works by male writers, so that books written by men and women ended up with an even gender ratio in each decade. (Because there are fewer works by women in the corpus, we also expanded each model to cover a twenty-year segment of the timeline.) Biographical information about the author is, as you might expect, hard to infer from a set of 54 words associated with a single character. 50% accuracy would be random, so a median accuracy of 57.4% will impress no one. But the point of this exercise is not really to model authorial identity. We're simply providing a point of comparison that may clarify the significance of changes in our character model.

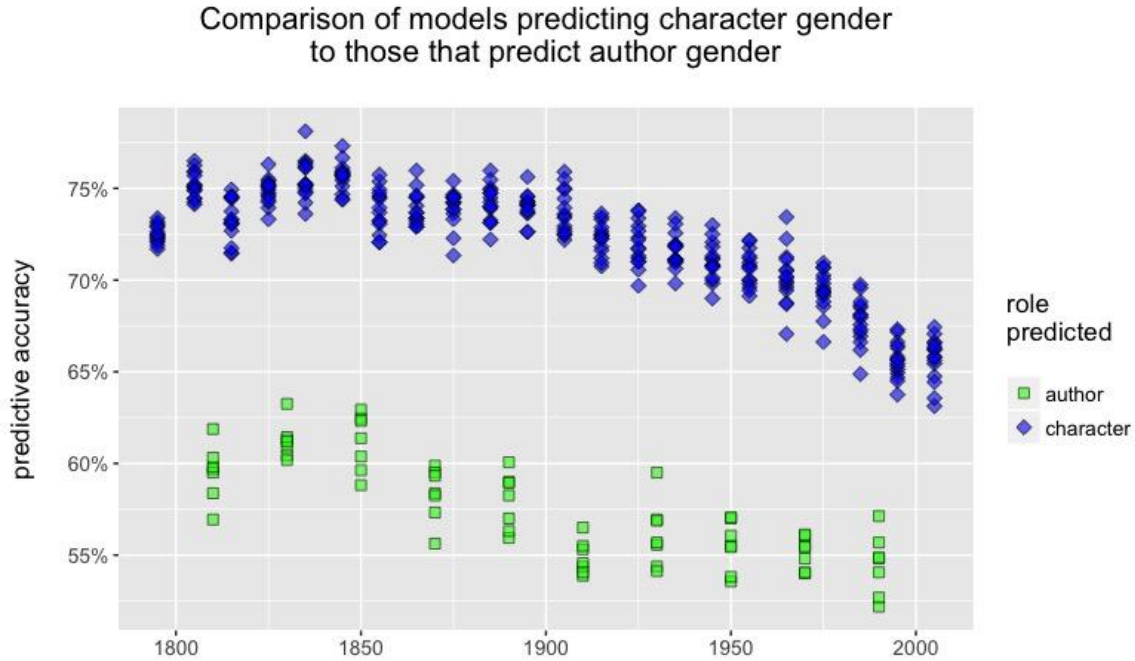


Figure 10. Models that use text from a single character to predict either the gender of the character, or of the author. Both sets of models use the 2200 most common words in a group of 1600 characters.

In this expanded context, it becomes clear that the changes in the character models are substantial. The two groups of models in Figure 10 are always widely separated, because of course description of a character speaks more explicitly about a fictional person than about the real person doing the writing. But by the twenty-first century it is sometimes (in exceptional 1600-character samples) as hard to infer the gender of characters as it had been to infer the gender of authors in the early nineteenth century. *Sometimes* is an important word in that sentence: the area of overlap is very small. To put this another way: fictional gender roles have weakened substantially by the twenty-first century. But they could get a lot weaker before it would be appropriate to describe them as faint or impalpable.

Another thing readers may notice in Figure 10 is that the decline in accuracy of author inference loosely tracks the decline in accuracy of character inference—or perhaps precedes it. We are not ready to draw many conclusions from this evidence: the apparent trend for authors depends on comparison of only four or five slices of time, and the trend can look significantly different with slightly different data or

parameters. But if the apparent pattern turns out to be durable, it would raise interesting questions about the relation between biographical and fictional gender roles. We won't have space to explore those questions fully here, but they might repay further study.

Leaving deep causal inferences for another day, we can at least attempt a rich description of trends in fiction. What were the implicit assumptions that made characters relatively easy to sort by gender around 1850, and why did the task of inference get harder? There are many ways to think about this question. You could approach the problem, for instance, by cracking open the models themselves and reasoning about coefficients. But in a complex model, coefficients can be tricky to interpret. Let's try a simpler approach. If we randomly select 10,000 words describing women, and subtract 10,000 words describing men, how much of a surplus (or deficit) will we have for each dictionary entry? In expectation, this is equivalent to simply taking the difference in a word's normalized frequency of use between men and women and scaling that difference to reflect the real number of additional (or fewer) occurrences we would expect to see in 10,000 words when moving from characters who are men to characters who are women.²⁰ This measure tends to emphasize common words, but for our purposes that's fine: we want to find changes that were big enough to make gender hard to infer. Here, for instance, are three common verbs.

²⁰ This is a form of "difference of proportions"; for more on this, and alternative metrics, see Burt L. Monroe et al. (2008), "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict," *Political Analysis*.

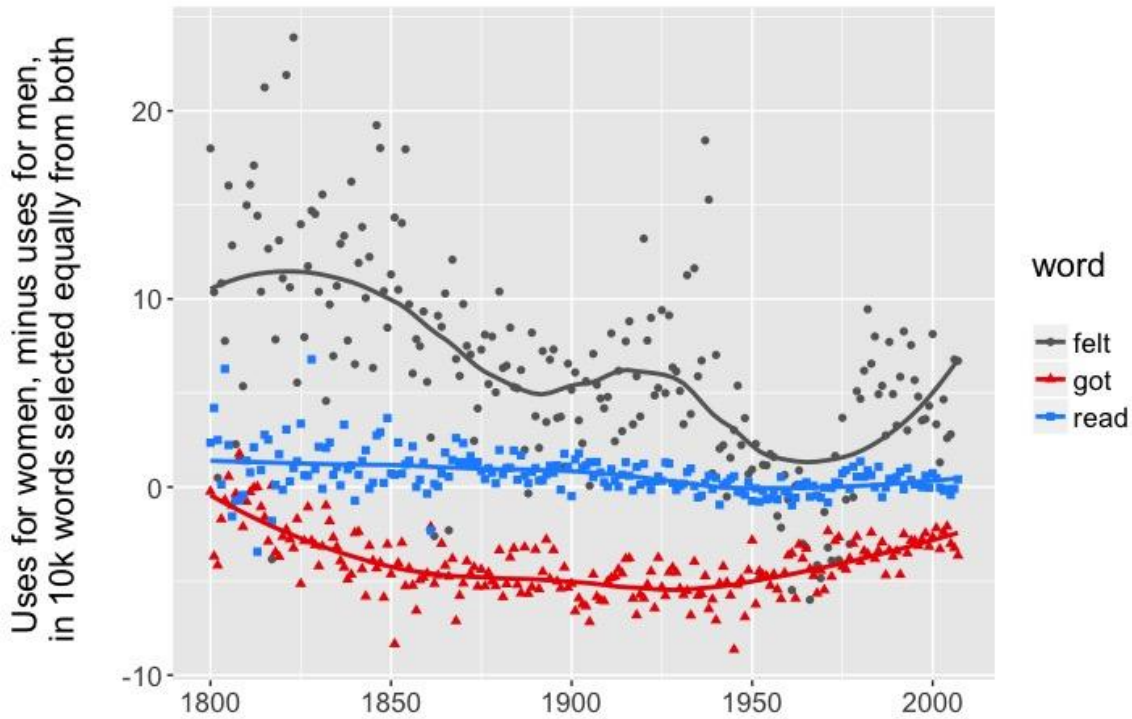


Figure 11. *The difference between a word's frequency in descriptions of women, and of men.*

“Reading” is an action that fictional men and women perform equally often, and that balance stays pretty equal across two centuries of fiction. But “got” is governed more often by a masculine subject, and “felt” by a feminine one. However, the surplus of feminine feeling declines as we move forward on the timeline, either because “felt” was used more often to describe men, or because it was used less often to describe everyone (this graph doesn't try to separate those changes). There's even a brief period in the 1960s when the word is used more often for men.

We're looking for words like “felt” — words that created a lot of gender differentiation early in this period, but less toward the end. Words with that kind of arc might explain why the language used to describe character became (on the whole) less gendered over time. Instead of choosing words at random, we can actually sort them to find words that move toward the central dividing line. When we do, we come up especially with a lot of things like this:

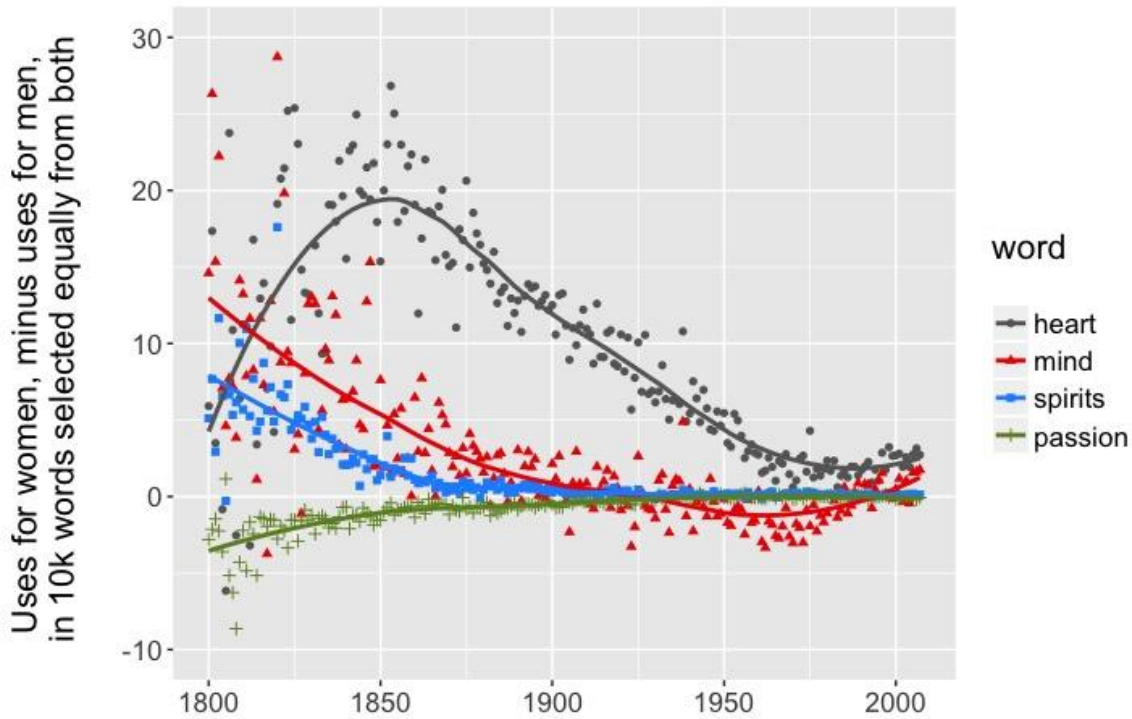


Figure 12. Words selected because they produce more differentiation early in the timeline than toward the end.

Language of thought and feeling, in general, was gendered feminine in the nineteenth century. This is true not only for *heart* (and *tears* and *sighs* and *smiles*), as we might expect, but for less emotive terms like *mind* and *spirits*. There are only a few subjective nouns ascribed more often to men; the primary one is *passion*, which is sometimes a nineteenth-century euphemism for lust. (If we are concerned about the difference between *mind* [verb] and *mind* [noun], BookNLP does allow us to separate parts of speech. In practice, the trend is so dominated by the noun that separation would make little difference.) Andrew Piper’s forthcoming work on character suggests a further twist to this evidence: he argues that the distinctive interiority of feminine characters was most marked in books written by women.²¹ We see some evidence to support his conclusion, although, for the particular words considered above, it appears

²¹ Andrew Piper, *Enumerations: Data and the Elements of Literature* (University of Chicago Press, forthcoming).

to be a question of degree. Books written by men and women agree in characterizing women more often with terms like “mind,” “spirits,” “heart,” and “soul”; women are just slightly more consistent in treating the terms as feminine.

The pattern we are seeing here is congruent with Nancy Armstrong's well-known thesis that subjectivity was to begin with “a female domain” in the novel: “It was at first only women,” she says, “who were defined in terms of their emotional natures.”²² For Armstrong this is not fundamentally a stereotype about women. She traces the gendering of subjectivity to a broader fantasy that social order can be guaranteed, without violence, by individual sentiment. But eighteenth-century and Romantic-era novelists often dramatize fantasies about political consent through stories about sexual consent that hinge on the transformation of a woman's feelings. In Armstrong's view, the point was not necessarily to deny the possibility of male emotion; the feelings of men just didn't have the same importance to this social fantasy. Only as we move toward the middle of the nineteenth century does psychological depth become equally important for male characters; in Armstrong's narrative, this shift is dramatized with a reading of Heathcliff in *Wuthering Heights*.

The evidence we are seeing also supports Armstrong's contention that the gendering of privacy and interiority was linked to a broader division between public and domestic spaces.

²² Nancy Armstrong, *Desire and Domestic Fiction: A Political History of the Novel* (New York: Oxford University Press, 1987), 4.

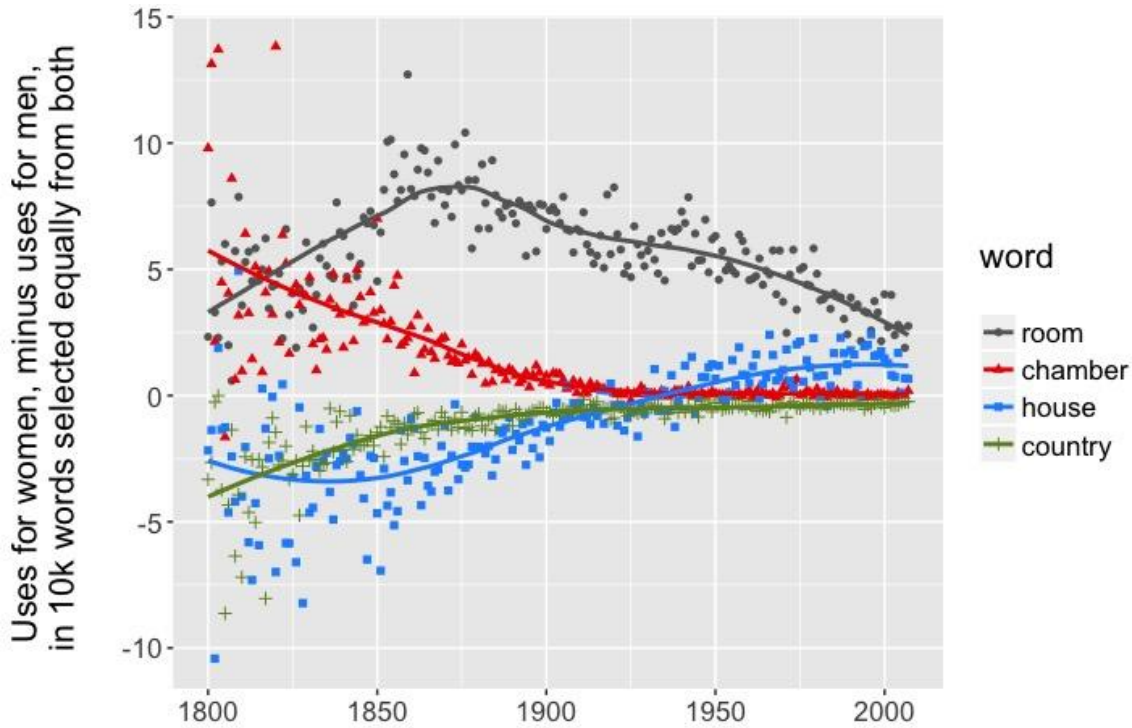


Figure 13. Spaces that are possessed by women, or by men, respectively.

In early-nineteenth-century novels, men have houses and countries. Women have private chambers and apartments inside the house, and once those words become old-fashioned, they have rooms. But this differentiation of spaces slowly declines across two centuries. By the later twentieth century, in fact, the *house* is associated slightly more often with women. (This could reflect the decline of coverture, allowing women to share legal ownership of a house—or perhaps also a growing sense that the *whole* suburban dwelling counts as private domestic space.)

These familiar changes go some distance toward explaining the growing blurriness of gender boundaries. But growing blurriness is not the only story you can tell with this evidence. In fact, there is far more going on in this data than we could completely describe in one article. (Readers are encouraged to explore the underlying data using [an interactive visualization](#).) Our thesis here has highlighted long-term trends, but a close observer will also notice fascinating little stories by the wayside. Some of these involve the creation of new forms of gender differentiation. For

instance, verbs of mirth become strongly gendered in the middle of the twentieth century.

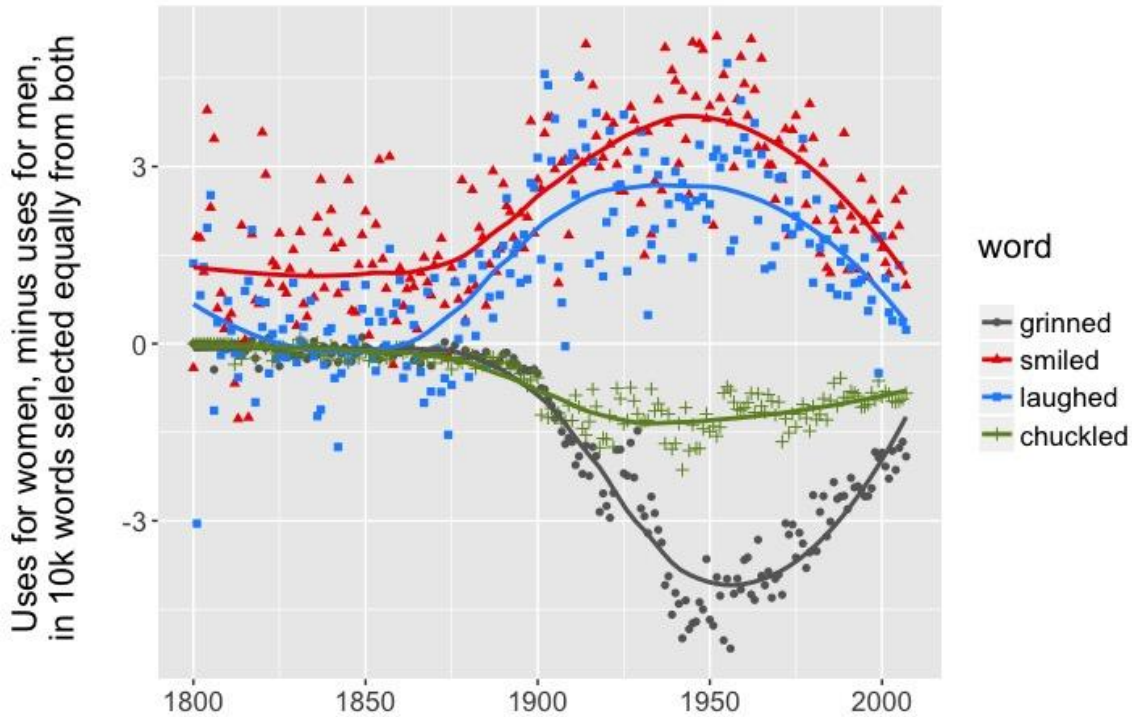


Figure 14. The mid-twentieth-century gendering of mirth.

Women smile and laugh, but mid-century men, apparently, can only grin and chuckle. At this stage of research, we cannot fully explain this odd phenomenon, but it is interesting that the feminization of the common words *smile* and *laugh* seems to have preceded the development of masculine alternatives—almost as though it became inappropriate for men to smile once women were doing it so much. This gendering of mirth peaks in the years before and after World War Two, and Raymond Chandler is a typical expression of its consequences for men. His male characters have a habit of grinning in an uneasy laconic way. “He grinned. His teeth had a freckled look,” in “Red Wind.” “He grinned. His dentist was tired of waiting for him,” in “The Pencil.”²³ Chandler's grins are commonly followed by a cynical chaser about the

²³ Raymond Chandler, *The World of Raymond Chandler: In His Own Words*, ed. Barry Day (New

character's appearance, to make clear that masculine humor is a thin veneer stretched over menace.

These expressions of gender can be periodized and dated because they are on their way out now. But other expressions of gender have continued to expand. In particular, description of the human body becomes steadily more important in fiction. Ryan Heuser and Long Le-Khac have shown how broad, steady, and dramatic a trend this is.²⁴ And as writers spent more time describing their characters physically, those physical descriptions also became more specifically gendered. A whole article could be written on this topic: jaws, hands, lips, and feet are all fascinating. But here are a few leading examples.

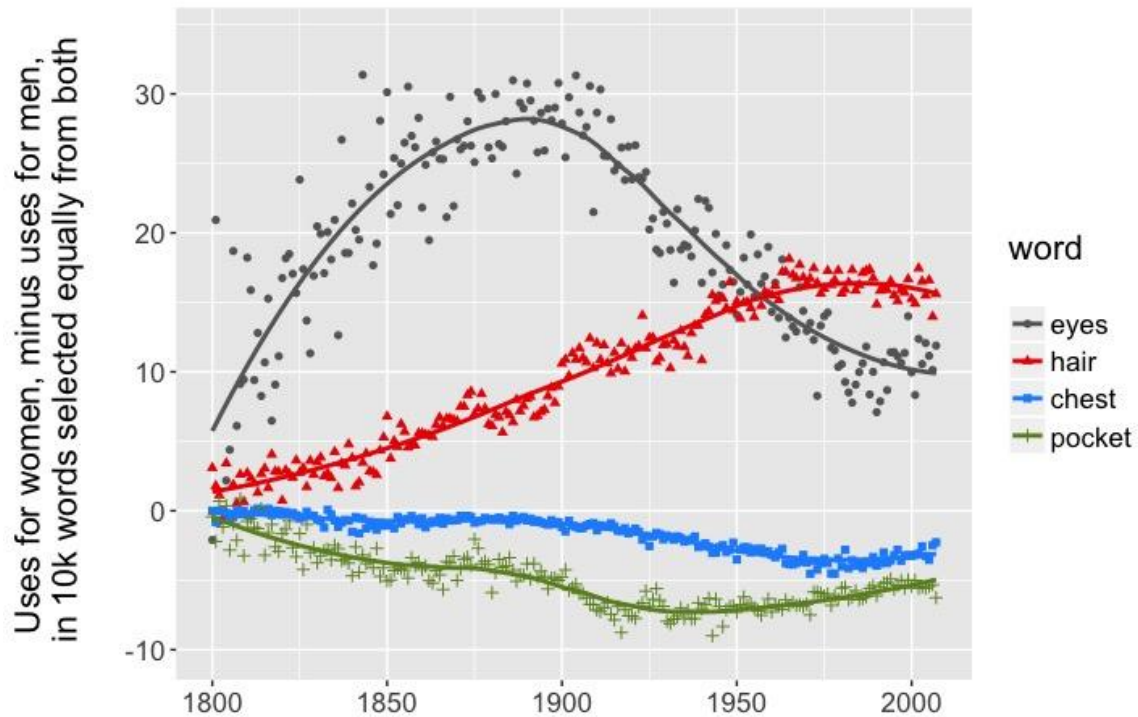


Figure 15. The rise of physical signs of gender.

York: Penguin Random House, 2014), 211.

²⁴ Ryan Heuser and Long Le-Khac, “A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method.” Stanford Literary Lab Pamphlet Series, 2012, <https://litlab.stanford.edu/LiteraryLabPamphlet4.pdf>.

There are actually several different waves of physical description. The top curve for *eyes* is emblematic of several other terms that peak in the late nineteenth century. Many of these involve the feminine head: *lips*, *eyes*, *face*, and *voice* are all associated with women. Then these powerful signs decline, making way for later waves of gendered physical description that become important in the twentieth century. Many of these involve the body or clothes. The masculine chest slowly becomes important, although bodily description is never quite as central for men as it is for women. (*Body* itself, in fact, is gendered feminine.) One of the main things men do have is a pocket; in the twentieth century they are constantly putting things in it. At present, hair is one of the most powerful signifiers of femininity.

So pulling all this together, what can we conclude? Some forms of gender differentiation (associated for instance with domestic space and subjectivity) are declining while other forms (associated for instance with the body and clothes) are on the rise. If you add them all together, we may be able to say generally that gender is less insistently marked by the end of the twentieth century than it was in the 1840s. But that slow increase in blurriness could be less important than the churn we have seen along the way: the rise and fall of different forms of gender differentiation. Although the opposition of *he* and *she* remains grammatically the same, gender is actually a different thing by 2007 than it had been in 1840.

Conclusion

This essay has considered several kinds of evidence that may seem to point in different directions. Fictional characters are implicitly gendered in ways that can be quite volatile. Tears and sighs matter in one period, chuckling and grinning in another. In fact, the volatility runs even deeper: gender as such seems to shape characterization more pervasively at certain points on the timeline (like the middle of the nineteenth century). On the other hand, certain aspects of fictive gender are surprisingly stable, especially when we back out to look at the behavior of authors. Men remain—on average, as a group—remarkably resistant to giving women more than a third of the character-space in their stories. And whatever happened to blur the

boundaries of gender as we move into the twentieth century, it doesn't seem to have been associated with greater emphasis on women as characters. On the contrary, their prominence declines across the same period.

We don't pretend to have fully explained these paradoxes. But there may be hints toward an explanation in *Edging Women Out*. For instance, if the proportion of novels written by women declined because men moved into a formerly feminine genre, while women were taking advantage of new opportunities elsewhere (for instance, writing nonfiction), then it would appear that genres themselves were becoming less strongly gendered. If that were true, it might make intuitive sense that gender roles inside fiction would be blurring at the same time. Indeed, it might not be a paradox at all that women writers left the genre where they had previously been segregated at the same time as gender roles were growing more flexible. This would seem paradoxical to literary critics only because we have a special professional interest in fiction, tempting us to view it as a barometer of equality generally—which it may not be at all. But that hypothesis opens a big question; to say that it will require further study is an understatement.

We have also deliberately left some theoretical questions open. Should changes in the expression of fictive gender be understood as a *convergence* between previously distant positions, as the *blurring* of a boundary to produce a spectrum, or as a *multiplication* of gender identities that made the binary opposition between masculine and feminine increasingly irrelevant to characters' plural roles? We have deferred that theoretical debate, in order to say simply that binary gender categories became less central to characterization.

Theoretical debate about gender can't be bracketed forever. Empirical research starts from theoretical assumptions, and produces theoretical implications. But this needn't be a purely circular process. Research can, after all, turn up evidence that wasn't implicit in the initial data model.²⁵ In this article, we took conventional

²⁵ Data models are admittedly more constraining for projects that aim to construct durable shared datasets. See Miriam Posner, "What's Next: The Radical, Unrealized Potential of Digital Humanities," *Miriam Posner's Blog*, July 27, 2015, <http://miriamposner.com/blog/whats-next-the-radical-unrealized-potential-of-digital-humanities/>.

binary roles as our starting-point, but found that they were in practice unstable. In one century, men had houses; in another, they had pockets and tense grins. In one century, women were characterized through their hearts; in another, through their hair. The overall strength of gender categorization also varied substantially from one century to the next. We have, in other words, measured the diachronic instability of gender categories.

The next kind of instability researchers might want to explore is perspectival. Whether we understand gender with Judith Butler as a performance, or (with Linda Martín Alcoff) as a real, although historically constructed, “position one occupies,” it is clear that gender is a relational category.²⁶ It is less a fact about the subject herself than about her relation to a social audience. But there can be, of course, more than one audience. Literary gender may be constructed differently in different genres, or in different parts of the literary field. Because predictive models are good at capturing implicit assumptions, they are well suited to teasing out these different perspectives.

For instance, how is femininity represented differently in books by women or by men? We have already mentioned, above, that models trained and tested on books by men find gender easier to predict. But what specific differences are created by authorial perspective? When we compare models closely, it turns out that women and men (surprise!) have substantial disagreements about gender. We only have room here for a brief and tentative discussion of this topic. But when we trained six models on different random samples of characters from 1800 to 1999, we found, for instance, that women writers consistently characterized their feminine characters with words like “spend,” “conscience,” “busy,” and “endeavoured,” which books by men had treated as masculine. In books by women, men are often the direct object of the verb “marry”; in books by men, women are. Different sexual identities, obviously, might introduce other kinds of perspectival variation here.²⁷

²⁶ Judith Butler, *Gender Trouble: Feminism and the Subversion of Identity* (New York: Routledge, 1990). Linda Martín Alcoff, *Visible Identities* (Oxford: Oxford University Press, 2006), 148.

²⁷ For a deeper exploration of these perspectival differences, in British and Czech fiction, see Anna Čermáková and Lenka Fárová, “His Eyes Narrowed — Her Eyes Downcast: Contrastive Corpus Analysis of Female and Male Writing,” *Linguistica Pragensia* 28.2 (2017): 7-34.

There is also a large area of agreement about gender. If we compare the predictions made by different models, those trained on books by one gender correlate at $r = 0.549$ with those trained on books by the other. But since men and women don't agree with each other quite as much as men do with other random samples of books by men ($r = 0.666$) or women with other samples of books by women ($r = 0.594$), it doesn't necessarily seem safe to conclude that we are all talking about the same categories. In short, quantitative methods could be used not only to trace changes over time, but to break the two categories we began with into four different relationships between observer and observed. Further research on sexuality, nationality, or genre could take us from four perspectives to twelve or twenty. (The grinning men in Chandler's *Big Sleep* are very different creatures from those in, say, Daphne du Maurier's *Rebecca*.)

At some point, of course, predictive models will reach their limits. If we want to discuss specific characters in *Rebecca*, the traditional methods of literary criticism will give us richer descriptive resources. A binary model of gender also has limits, even as a heuristic premise. If we want to bracket gender binaries altogether, we can do that at the very outset of a research project. Scholars applying computational methods to social media have used a flexible Butlerian theory of gender to good effect.²⁸ But if we are interested in teasing out the perspectival and historical variations in apparently stable categories, predictive models can be very useful. This essay has come nowhere near exhausting their potential; these methods have a lot of room to run, and they could eventually have broad theoretical implications.

²⁸ David Bamman, Jacob Eisenstein, and Tyler Schnoebelen, "Gender Identity and Lexical Variation in Social Media," *Journal of Sociolinguistics* 18.2 (2014): 135-60.