

How Quickly Do Literary Standards Change?

Ted Underwood and Jordan Sellers¹

Our discipline's long struggle over pedagogical canons taught us to imagine that literary-historical inquiry naturally begins by grounding representation on a single, normatively balanced list of works.² That list could be the canon itself, if we believed it contained the most important books — or it could be an alternate list, expanded and rebalanced according to better criteria of significance. Or perhaps it could be a list of every title ever published, if anyone claimed to have such a thing.

This attempt to locate a correct representative foundation for our field made more sense as a struggle over cultural authority than it does as a starting-place for historical inquiry. It might actually be a bad idea for literary historians to agree on a single synecdoche for the past. Different questions require different comparisons — across genres, or demographic groups, or forms of literary prestige. Even if it were possible to create a single list balanced across all these dimensions at once, it's not clear that it would be useful. Social scientists don't begin, after all, by agreeing on a list of representative individuals. They begin by framing specific questions and selecting samples to address them.

Distant readers have likewise based their theory of historical representation on sampling.³ Every literary-historical sample is limited by a particular method of selection; no single sample provides a complete picture of the past. But we can still learn a lot by comparing samples. For instance, Mark Algee-Hewitt and Mark McGurl have recently compared bestsellers to novels selected by different groups of critics, to discover how different forms of literary significance overlapped in the 20th century. (It turns out that they mostly didn't.)⁴ In this sort of inquiry, the advantage of a large collection is not that we know it to be complete or correctly balanced, but that

¹ The design of this project was probably influenced by “Canon/Archive,” a work in progress involving several universities (Sarah Allison, Marissa Gemma, Mark Algee-Hewitt, Ryan Heuser, Franco Moretti, and Hannah Walser). It has also benefited from conversations with colleagues in Provo and Seattle.

² John Guillory, “Canon, Syllabus, List: A Note on the Pedagogic Imaginary,” *Transition* 52 (1991): 36-54.

³ See Franco Moretti, “Conjectures on World Literature,” *New Left Review* 1 (2000): n21.

⁴ Mark Algee-Hewitt and Mark McGurl, “Between Canon and Corpus: Six Perspectives on 20th-Century Novels.” Stanford Literary Lab Pamphlet Series (Jan 2015). http://litlab.stanford.edu/?page_id=255

it creates room for different sampling strategies, giving us contrastive leverage on a wide range of questions. If we're studying the history of literary production, we may need every title we can find. For other questions, we may ground our research on a list circumscribed by some measure of literary significance. But if we can find the right samples to contrast, it may even be possible to dig beneath that foundation, and explain how literary significance itself was created and transformed.

That's what we'll attempt in the pages that follow. We have created two samples of poetry and fiction across the century 1820–1919: one drawn from reviews in prominent Anglo-American periodicals that reviewed literature selectively, the other drawn at random from HathiTrust Digital Library, which contains at least 146,000 English-language volumes of poetry and fiction in this period, many relatively obscure.⁵ The contrast between the provenance of these two samples allows us to frame synchronic models of literary prestige for particular periods. But we're also interested in a diachronic problem: how quickly did the standards governing prestige change, and how were those directions of change related to the synchronic axis of distinction?

We've found some things we expected, and others we didn't. But we began our inquiry, at any rate, with an explicit hypothesis. We expected that a widely-discussed "great divide" between elite literary culture and the rest of the literary field would become increasingly visible as we proceeded through the late nineteenth century and into the twentieth.⁶ We proposed to trace the emergence of that divide with predictive modeling. As different literary styles specialized to address different reading audiences, we believed it would get easier and easier to predict whether a given volume had been reviewed in a selective venue, merely by looking at the words in the text itself.

We found something different. It certainly is possible to use the diction of a literary text to predict whether it was reviewed in a prestigious venue, but we haven't found much evidence that this differentiation of styles emerged when we thought it

⁵ Most of the volumes in both samples were published in Great Britain, Ireland, Canada, or the United States.

⁶ Andreas Huyssen, *After the Great Divide: Modernism, Mass Culture, Postmodernism* (Bloomington: Indiana University Press, 1986), viii.

might (the late nineteenth century). Instead, we've found a social divide that seems to have been present from the very beginning of the period studied. But in pursuing this hypothesis, we've also stumbled on a number of other patterns that speak to broader questions about the pace and direction of literary-historical change. The inquiry has turned out to be so productive, in fact, that this article will only be able to describe one half of it — the part concerned with poetry.

The plan of the experiment

Before going further, we should spell out which aspects of this experiment are assumptions, and which hypotheses. It may seem strange to assume that we can map literary prestige by asking whether volumes were reviewed in a particular set of periodicals. How could we know which venues counted as prestigious? For that matter, how could any model of prestige hope to cover a whole century? This project covers everyone from William Wordsworth to Amy Lowell, with Spasmodics, Fireside Poets, and pre-Raphaelites in between. Reviews will be drawn from fourteen different periodicals on both sides of the Atlantic. It seems unlikely that all these different writers and reviewers could have shared a single set of aesthetic standards.

We don't assume that they did. If we choose venues of review that turn out to have no selection standards in common, the works in our reviewed sample won't share many characteristics, and it will be hard to distinguish them from the random sample. Likewise, if it turns out that the nature of literary prestige changed completely over the course of this century, we will discover that it's impossible to model the century as a single unit. In fact, since it seemed likely that literary standards had been too volatile to generalize about a whole century at once, we initially set out to model four twenty-year periods between 1840 and 1919. But the notion that standards would remain stable even for twenty years was not an assumption we made: it was a hypothesis we set out to test.

The main assumption we will make is that being reviewed at all indicates a sort of literary distinction — even if the review is harsh.⁷ We could have chosen to model distinction instead as a boundary between positive and negative reviews. That is, after all, how literary historians have typically understood this question. It’s an approach that makes sense if you’re interested in gradations of approval between well-known writers, but it leaves out a mass of works that were rarely reviewed at length in selective venues. This is where digital libraries are giving us new leverage, making it possible to see large numbers of things that never made it into our literary histories, and thereby revealing a social boundary that would be invisible from within our received narrative. So, although we recorded reviewers’ sentiments when they were clear, we placed more emphasis on the mere appearance of a title in certain venues.

To make this strategy work, we needed to focus on periodicals that were selective about what they did review — usually quarterly, monthly, or fortnightly reviews, rather than weeklies. We created an initial list by crowdsourcing the question among friends who are scholars of nineteenth or early-twentieth-century literature.⁸ Then we winnowed that list by choosing journals that seemed especially selective in their literary reviewing. For instance, *The Athenaeum* was influential but reviewed so many novels that it’s not a sign of great distinction to be included there. Journals like *The Fortnightly Review*, with broadly intellectual ambitions, covered new fiction and poetry less often. Eminently good or eminently bad, literature reviewed there was at least marked as important.

This is the list of titles we chose. For each title, we have listed the earliest and latest publication date for volumes we sampled from it, and the number of volumes of poetry used in the present study. (Although 1820 was the earliest year we sampled reviews, dates of publication can be earlier. Note also that in a small number of cases

⁷ We don’t assume that it’s a distinction conferred *by the reviewer*. It’s likely that many periodicals decided what to review by looking at particular publishers’ lists, or by “puffing” their own publisher. Nicholas Mason, *Literary Advertising and the Shaping of British Romanticism* (Baltimore: Johns Hopkins University Press, 2013).

⁸ Thanks are due especially to Nina Baym, Ryan Cordell, Eleanor Courtemanche, Jeff Drouin, Andrew Gaedtker, Lauren Goodlad, Matthew Hart, Deanna Kreisel, Bethany Nowvickie, Anthony Mandal, Bruce Michelson, Justine Murison, and Roger Whitson.

— about 5% of volumes sampled — we used a work’s publication in a periodical rather than appearance in a review as a proxy for editorial judgment.⁹)

<i>The Atlantic</i>	122	1845–1905
<i>Blackwood’s Magazine</i>	16	1838–1896
<i>Contemporary Review</i>	2	1877–1878
<i>The Egoist</i>	17	1912–1918
<i>Edinburgh Review</i>	36	1819–1856
<i>Fortnightly Review</i>	60	1863–1917
<i>Graham’s Magazine</i>	19	1827–1855
<i>Macmillan’s Magazine</i>	2	1881–1881
<i>Poetry: A Magazine of Verse</i>	32	1910–1916
<i>Quarterly Review</i>	19	1816–1851
<i>The Savoy</i>	2	1896–1897
<i>Tait’s Edinburgh Magazine</i> ¹⁰	64	1840–1876
<i>The New Age</i>	2	1907–1908
<i>Westminster Review</i>	29	1828–1867
<i>The Yellow Book</i>	2	1893–1895

We also needed a sample that would be likely to contain books reviewed less often. To hand-select books that were never reviewed at all would have been challenging, and might perhaps have produced a narrow group of works. We decided it was more straightforward simply to select works at random from a very large pool (while excluding authors already in our reviewed sample). This is where digital libraries shine. We worked with HathiTrust, which contains the aggregated collections of large public and university libraries; in the period we’re considering (1820–1919), that gave us a collection of roughly 758,400 books in English, of which about 53,200 include significant amounts of poetry. This doesn’t exhaustively cover print culture; it’s still a sample, with particular selection biases.¹¹ But it’s a sample much broader than the social range covered by elite periodicals, and what we needed in this study was

⁹ This was a debatable choice. We did it in order to include influential publications like *The Yellow Book* that were not strictly speaking reviews. However, the sample is still predominantly shaped by reviewing practices. Methods of sampling are described more fully in a methodological appendix.

¹⁰ In this article, we won’t actually use the reviews sampled from *Tait’s Magazine* (as we explain in the methodological appendix, we concluded that it had been less selective than the other periodicals on this list).

¹¹ Our estimates suggest that there are between 5,000 and 10,000 distinct poetry titles in this collection of 53,200 volumes. (Some are translations or works originally published before 1820.) This may represent more than half of the titles that were printed. Allen Riddell’s estimates on early-nineteenth-century fiction suggest that HathiTrust has about 58% of titles recorded in standard bibliographies. Our work on fiction suggests that it also has many titles left out of bibliographies. Allen B. Riddell, “Where are the novels?” 14 February 2012, <https://ariddell.org/where-are-the-novels.html>.

contrast. Finding literary works in a large digital library is a task in itself; in fact half the time we spent on this project was spent locating the pages that actually contained poetry and fiction in nearly a million volumes.¹² But once you do have a map on that scale, random sampling is likely to turn up works that are very different from those that got reviewed. We gathered 360 reviewed and 360 random volumes, distributed in a similar way over the timeline.

A model of reception

The models we discuss in the pages that follow represent the world by making testable predictions. For instance, looking only at the words in a volume, the model will predict which sample it came from: was it found in a review, or sampled randomly from a digital library? The point isn't literally to predict events that already happened, but to discover whether there's any relationship between poetic language and reception. We test the model's account of that relationship by asking it to make predictions about volumes it hasn't yet seen, in order to find out whether the model has learned anything generalizable, or just memorized a particular set of examples.¹³

All the model knows about each book are the relative frequencies of the words it contains. This representation is different from human readers' sequential engagement with language, and the uninitiated often assume that it shouldn't reveal much. But as literary critics, we're familiar with the notion that words signify on multiple levels. Perhaps it won't surprise us that the choice of a single word can reflect a text's genre and tone and stance toward an imagined audience, as well as its explicit themes. In any case, computational analysis of text has relied on word frequencies because they do turn out to register a lot of different things.¹⁴

¹² Ted Underwood, "Page-Level Genre Metadata for English-Language Volumes in HathiTrust, 1700-1922," figshare, December 2014, <http://dx.doi.org/10.6084/m9.figshare.1279201>.

¹³ The advantages of predictive models are perhaps best explained in Leo Breiman, "Statistical Modeling: The Two Cultures," *Statistical Science* 16.3(2001): 199-231.

¹⁴ "For many tasks, words and word combinations provide all the representational machinery we need to learn from text." Alon Halevy, Peter Norvig, and Fernando Pereira, "The Unreasonable Effectiveness of Data," *IEEE Intelligent Systems*, 2009. <http://static.googleusercontent.com/media/research.google.com/en/us/pubs/archive/35179.pdf>

At the grittiest mathematical level, the model we’ll discuss is just an equation that translates word frequencies into a probability that a particular volume came from the reviewed sample. We create the equation by showing a regression algorithm¹⁵ the volumes from all authors (except one) in both samples; the algorithm assigns each word a positive or negative weight in an effort to separate the samples as reliably as possible.¹⁶ That process “trains” a model. When we show the model a new volume, by the author it hasn’t yet seen, it can make a prediction using the weights assigned to different words. To avoid circularity, it’s important that the model doesn’t make predictions about one of an author’s books using information about others.¹⁷ So we actually have to train 636 slightly different models — each one excluding books by a different author. However, since each pair of models shares more than 99% of their evidence, it’s a fair simplification to characterize this group collectively as a single model of the literary field. If we plot all the volumes in a space where the y axis is defined by the model’s degree of confidence that a volume came from the reviewed set, this is the pattern we find:¹⁸

¹⁵ Logistic regression has been around since 1944; the version we use is better able to handle lots of variables because it adds “regularization,” but this is not a recently-invented black box. Joseph Berkson, “Application of the Logistic function to Bio-assay,” *Journal of the American Statistical Association* 39 (1944): 357–65. Robert Tibshirani, “Regression Selection and Shrinkage via the Lasso,” *Journal of the Royal Statistical Society, Series B* 58 (1996): 267–288.

¹⁶ We used the logistic regression functions from [Scikit-learn: Machine Learning in Python](#), Pedregosa *et al.*, *Journal of Machine Learning Research* 12 (2011): 2825-2830. And our code is available at github.com/tedunderwood/paceofchange, so our models can be replicated or adjusted to make different hermeneutic choices.

¹⁷ D. Sculley and Bradley M. Pasanek, “Meaning and Mining: The Impact of Implicit Assumptions in Data-Mining for the Humanities,” *Literary and Linguistic Computing* 23.4 (2008): 409-424.

¹⁸ Hadley Wickham, *ggplot2: elegant graphics for data analysis* (New York: Springer, 2009).

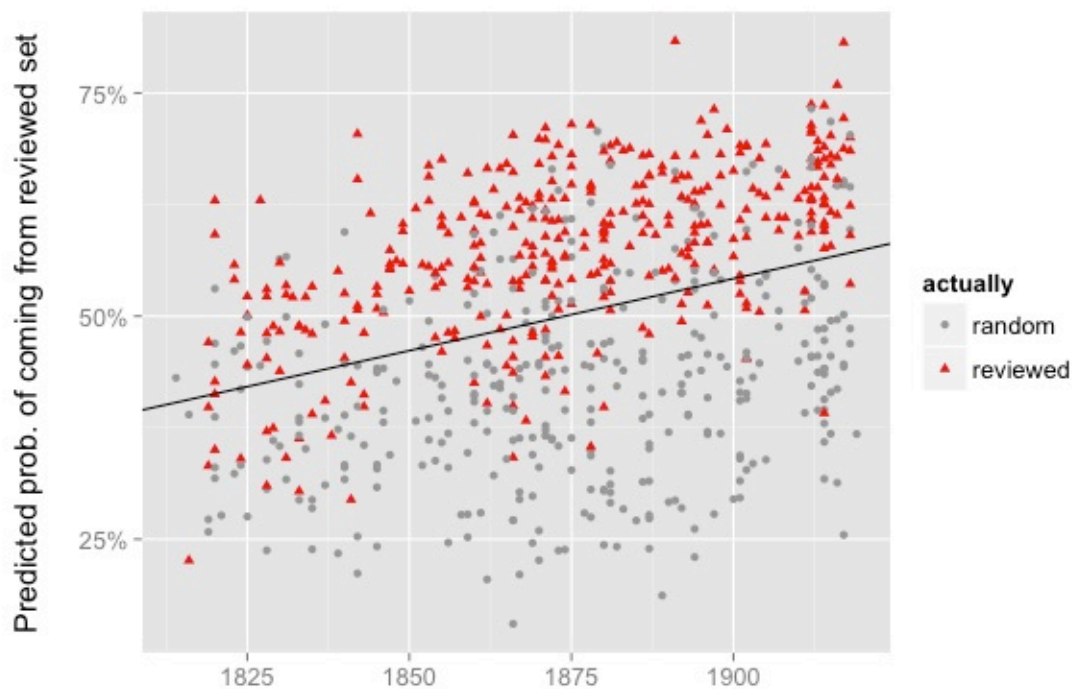


Figure 1. Predicted probabilities that volumes come from the reviewed set.

There’s a lot of information here to unpack. But just glancing at this graph you can see, first of all, that our model does a pretty good job of sorting reviewed from random works — and, somewhat to our surprise, can do so for the whole century at once. This is probably the simplest and biggest discovery we’ve made: the verbal differences between prominent and obscure authors turn out to be relatively consistent over long spans of time.¹⁹

But how consistent, exactly? Normally we would evaluate a model of this kind using the 50% line in the middle of the y axis; the model predicts that everything above that divide probably comes from our “reviewed” sample. Evaluated in that simple way, the model would perform moderately well; it’s right 77.5% of the time. But we can get better accuracy by acknowledging the odd fact that the whole collection drifts upward as historical time passes. If we consider publication date as a

¹⁹ A team of researchers at Stanford (including Ryan Heuser and Mark Algee-Hewitt) discovered much the same thing perhaps a year before we did — as they unfortunately often do. The Canon/Archive project, involving scholars from multiple universities, will cast new light on this pattern.

factor, and use the slanted black line to divide the dataset, the model will be 79.2% accurate.²⁰ Technically, the upward drift we're acknowledging is an error in the model. Volumes are not really "more likely to be reviewed" just because they were published later. But this is an error of an interesting kind, because the upward drift suggests that the criteria of literary prestige are bound up somehow with historical change. The model doesn't know when these volumes were published: they drift upward simply because the words that were more common in reviewed volumes across this period are also words that tended to become more common in *all* volumes by the end of the period.

That's interesting. But before drawing inferences about historical change, we should more fully assimilate the significance of the fact that we're able to divide reviewed from randomly-selected volumes with 79.2% accuracy. How good is that, and how good would it have to be, before we would call it meaningful?

We should begin by asking why anyone would expect this sort of prediction to work at all. Algorithms can often infer genre with greater than 90% accuracy, which is why we rely on spam filters to detect advertisements. But the social provenance of a document is usually harder to detect.²¹ And in the case of reception, we're making inferences about a social event that would have happened to a text only *after* it was written. Decisions about that event could have been made by hundreds of different people on opposite sides of the Atlantic, guided by standards of poetic significance (or other agendas) that were far from explicit and often a subject of fierce disagreement.

So it was entirely possible that there wouldn't be any pattern here to detect. The point of trying to predict reception isn't to automate a familiar task, like spam filtering — it's to discover whether reception was actually governed by thematic or stylistic regularities at all. The answer to that question could be interesting, even if it turned out to be "no." (Accuracy close to 50% would amount to a no, because in an evenly-divided dataset that's just guessing). A model of reception could also be

²⁰ We create "the slanted black line" very simply by running linear regression on the whole dataset to identify a central trend line. More sophisticated approaches are possible but unnecessary.

²¹ Nationality, for instance, may be more difficult to infer than reception; an analogous model identifies American and British poets with around 77% accuracy. (Words like "theatre" might be an easy clue, except that we routinely normalize spelling for all the texts we use.) Gender is even harder to infer.

interesting even if it only turned out to be stable across short periods — which seemed likely. If prominent writers became prominent by occupying the leading edge of a rapidly-moving wave, they might share different habits in different periods. Perhaps in the 1820s prominent poets would be united by gloomy Byronism, in the 1850s by an interest in history, and in the 1890s, by the word “mauve.” As for the randomly-selected volumes, who knows? They might only share a tendency to trail thirty years behind the wave. If poetic distinction had been created by this sort of volatile boundary between avant- and arrière-gardes, we would have needed to train different models to characterize the logic of prestige in different periods. And since received literary histories strongly suggest that one or more poetic revolutions intervene between 1820 and 1919, this was the first approach we tried.²²

In practice, however, we found that we could separate the samples most accurately by treating the whole century as a single unit, organized by a single set of standards.²³ In other words, you can use the same list of prestigious or banal words to distinguish Lord Byron’s *Prophecy of Dante* from less prominent works in 1819, and to distinguish Christina Rossetti’s *The Prince’s Progress* from less prominent works in 1866, and to distinguish T. S. Eliot’s *Prufrock and Other Observations* from less prominent things published in 1917. If a single list of words can predict literary prestige across that distance, some aspect of reception must be more stable than we anticipated. Reviewers may have disagreed about politics and about the merits of particular books, but they seem to have shared a loose, durable consensus at least about the outer boundary of distinction — the question, What kind of writing is even worthy of notice?

²² Scholars propose different dates, but almost everyone agrees that a “revolution” happened somewhere in this period. W.B. Yeats dated the “revolt against Victorianism” and against “the poetical diction of everybody” to the 1890s. See discussion in Richard Fallis, “Yeats and the Reinterpretation of Victorian Poetry,” *Victorian Poetry* 14.2 (1976): 89-100. Americanists sometimes identify “poetic revolution” with Whitman; the editors of the *NAEL* locate it slightly later. “The years leading up to World War I saw the start of a poetic revolution.” Stephen Greenblatt, et. al., *Norton Anthology of English Literature 8th Edition*, vol 2 (New York: WW Norton, 2006), 1834.

²³ For instance, models trained on two halves of the timeline were not collectively more accurate than a model of the whole thing. This remains true even when we randomly sample the century-level data to ensure that it’s working with the same number of volumes as the short-term models.

The pace of change

Distant readers have been trying to understand change for sixteen years now, and some patterns are starting to emerge. This article is in some ways a continuation of Franco Moretti's "Slaughterhouse of Literature." Both articles contrast prominent and obscure works to discover a system of differences that defines literary success. The authors of both articles are also taken off guard by the same aspect of their results. In "Slaughterhouse," Moretti set out to discover changes in the logic of plot across a single decade of detective fiction — and ended by concluding that those trends were probably diffused across a longer timeline.²⁴ Here we initially set out to model aesthetic preferences in four twenty-year periods, and found that it made more sense to model a century as a single unit. (We pushed the start date of our study back from 1840 to 1820 only after discovering that reception changed more slowly than we had expected.) In short, it's beginning to look like our received narratives of literary history lead us generally to overestimate the pace of change.

But a claim this large deserves a lot of skeptical testing. Predictive models are an unfamiliar kind of evidence. For instance, the model represented above asks a computer to find a linguistic definition of poetic prominence that can explain a whole century at once. It's significant that it succeeds, but that's not quite what we ordinarily mean by evidence of historical continuity, because the modeling process is actively trying to find an explanation that will cover this whole period. A more intuitive way to assess change might be to train models of reception on different segments of the timeline and then compare them. For instance, we could train a model only on volumes from one quarter-century, but ask it to make predictions about the whole century. That actually works: models trained only on a quarter-century of the evidence are still right (on average) about 76.8% of the volumes in the whole dataset. That's already strong evidence for continuity. If we want to understand the contours of change in more detail, we can compare the exact probabilities predicted by these models. It turns out that models trained on adjacent 25-year periods make very similar predictions; their predictions about the whole 720-volume collection correlate

²⁴ Franco Moretti, "The Slaughterhouse of Literature," *MLQ* 61.1 (2000): 207-27.

on average at $r = 0.933$. The two models on opposite ends of the century, whose midpoints are separated by 75 years, are least similar, $r = 0.779$. But that still means they share 67.7% of their variance.²⁵ That gives us a rough and ready answer to the question in our title: How quickly do literary standards change? At the boundary we're modeling (getting reviewed in selective periodicals), they change quite slowly. None of these models can explain reception perfectly, because reception is shaped by all kinds of social factors, and accidents, not legible in the text. But a significant chunk of poetic reception can be explained by the text itself (you can make predictions that will be right almost 80% of the time), and that aspect of poetic reception remained mostly stable across a century.

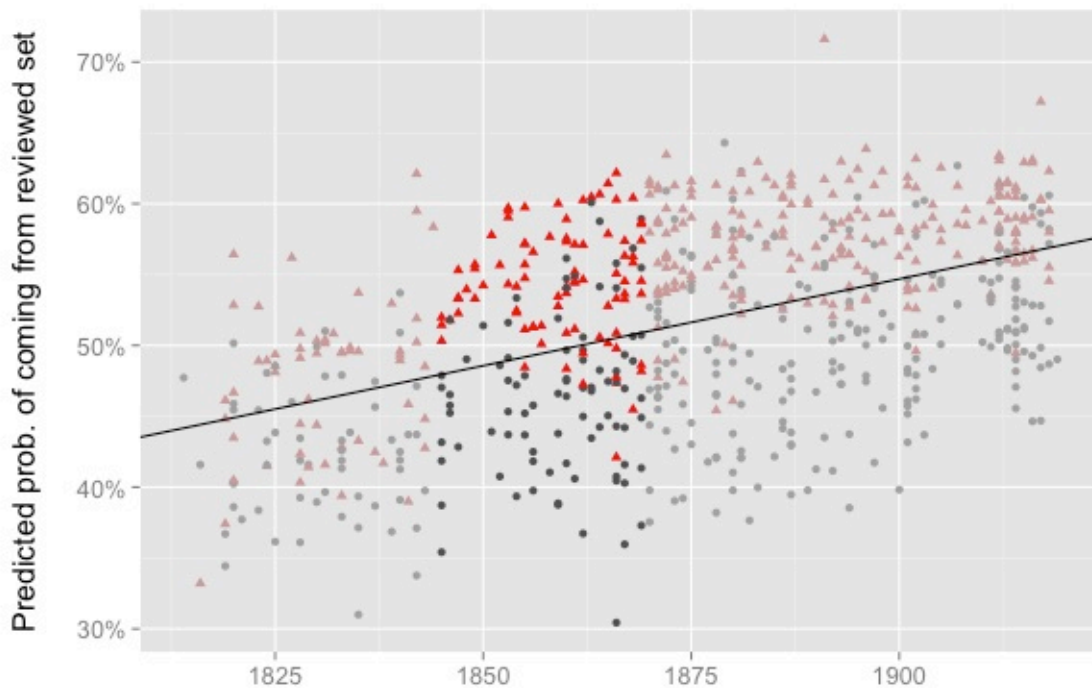


Figure 2. A model based only on volumes 1845-1869 makes predictions about the whole century.

The logic of poetic distinction

Since the canonical literary tradition seems too diverse to produce this kind of stable boundary, we suspected at first that the source of stability must be located in the

²⁵ We're taking the correlation between two models' predictions over the whole 720-volume dataset, and then squaring it to get proportion of variance explained.

random sample. The volumes of poetry we don't usually read must be united by some obvious feature: maybe they're all religious? Or just all blatantly awful? One way to test this hypothesis was to ask whether human beings would find it equally easy to identify the provenance of these texts. So we presented random pages from both samples to graduate students and professors who study nineteenth or early-twentieth-century literature, and asked them to guess whether each page had been selected from reviewed volumes or randomly sampled from a library. We told them when each volume had been published, and let them know whether each guess was right or wrong as they went along. Trained readers were right 64% of the time. It's not an apples-to-apples comparison with our model (which gets to read whole books, not single pages), but it does tell us at least that there's nothing screamingly obvious to human readers about the difference between these samples.

So how is a statistical model able to be right 79.2% of the time? The model we're training here uses 3200 variables — the frequencies of the 3200 words most common in the collection. A model this complex can encode a lot of information about a social boundary. But the complexity also creates room for a definition of literary prestige that can be about a lot of intersecting things at once; it's not required to map onto any single idea, and it's unlikely that we'll be able to explain the model completely here.

In practice, though, it's easy to sketch at least a few of the strongest patterns. We can see what the model likes about specific passages, for instance, by looking at the words that matter most. Each of the 3200 words most common in the collection has some influence on the model's prediction, but in many cases that influence is slight. Here we've only colored words red if the model expects them to markedly increase a poem's likelihood of being reviewed.²⁶ We've rendered words in blue if the model expects them to markedly decrease that likelihood. We'll start with the

²⁶ Interpreting statistical models raises methodological questions too involved to discuss here. See the note on "training and interpreting models" in the methodological appendix; the short version is that we feel confident about the patterns traced here, but don't intend to suggest that these are the only significant patterns.

conclusion of Christina Rossetti’s “Echo,” which the model sees as likely to be reviewed:

Yet come to me in **dreams**, that I may live
My very life again tho’ **cold** in **death**:
Come **back** to me in **dreams**, that I may **give**
Pulse for pulse, **breath** for **breath**:
Speak low, lean low,
As long ago, my love, how long ago.²⁷

One can detect traces of the advice that writing teachers still give: “use definite, specific, concrete language.”²⁸ The model loves “breath for breath”; it loves speaking and leaning low. By contrast, the abstract reflection on time in the last line doesn’t move the needle. If it were actually judging poems, the model would be wrong about that line, by the way: the dissolution of imagined immediacy into painful memory at the end is beautiful in context, and it’s apt that a poem called “Echo” ends with repetition. But the model isn’t judging quality. We should imagine it not as a critic, but as a literary agent offering broad advice about the general correlation between diction and reception.

Even in that capacity, it’s not quite consistent about recommending concrete language. It likes certain abstractions too, such as “dreams” and “death” — although, perversely enough, not “live” or “life.” It likes “low,” but not “high”; “hate,” but not “love”; “bitter,” but not “sweet.” In fact, we may as well come out and admit that this model is happiest when poems are a bit desolate. “Shuddering,” “blind,” “hollow,” and “blank” are some of its favorite words. It has an allergy to “kindness” and “valour.” It doesn’t even like “homes.” We can see why if we look at the volumes at the very bottom of its list — the ones it’s rightly confident will never be reviewed. Many of these have some inspirational or hortatory purpose; they’re about equally divided between religious and political topics. What they have in common is a tendency to affirm community in an abstract positive way. In *Memorial or Decoration Day* (1891), for instance, George Loomis invokes “those who battled for these **homes** of **ours**, / And **precious** blood on **Freedom’s altar** shed.” By no means all the volumes in

²⁷ Christina Rossetti, “Echo,” *Goblin Market and Other Poems*, 2nd ed. (London: Macmillan, 1865), 81.

²⁸ William Strunk and E. B. White, *The Elements of Style* (New York: Penguin, 2007), 37.

the “random” set are this sentimental, but there are enough (thoroughly obscure) examples of this style to make the model wary, not only of positive abstractions, but of the first-person plural in general.²⁹

This isn’t necessarily to suggest that political or religious themes prevented a writer from being reviewed. The model agrees with nineteenth-century opinion in rating Christina Rossetti very highly, although some of her verse is devotional. It also places Margaret Widdemer’s volume *The Factories* (1915) near the top of its list. But notice that Widdemer’s political rhetoric is neither abstract nor uplifting:

I have robbed my sister of the lips against her breast,
(For a coin, for the weaving of my children's lace and lawn),
Feet that pace beside the loom, hands that cannot rest—
How can she know motherhood, whose strength is gone?³⁰

Not every reader will like this rhetoric: constantly translating metonymic images back into ideas can be a bit exhausting. But Widdemer was widely reviewed, and indeed won a Pulitzer prize. So, these are the broad patterns that leap out immediately from a model of poetic reception 1820-1919: a preference for concrete language and a relatively dark tone (or at least, not a sentimentally uplifting one). As we’ve hinted, though, a complex statistical model resists summary; there are always other details beneath the surface. For instance, although the model generally loves gloom, it does roll its eyes at heavy-handed signals like “bleak” and “dire.”³¹ It’s also worth noting that feminine pronouns make a poem more likely to be reviewed. “She” and “her” aren’t red above because individual occurrences don’t have a huge effect, but if they’re repeated often, they add up.³²

²⁹ June Howard remarks that “in postbellum America, the literary was often defined *against* sentimentality.” Our study of reception supports that claim, while broadening it temporally and spatially. June Howard, “What is Sentimentality?” *American Literary History* 11.1 (1999): 73.

³⁰ Margaret Widdemer, *The Factories, With Other lyrics* (Philadelphia: John C. Winston, 1915), 10.

³¹ Interpreting a model with 3200 variables is not a rigorously experimental process. We’re using personification to acknowledge openly that some projection is inevitable.

³² A lot more could be said about the implicit gendering of poetry in this period. See Carol Christ, “The Feminine Subject in Victorian Poetry,” *ELH* 54.2 (1987): 385-401.

Another way to understand the model is to look at odd cases: for instance, writers whose careers cover a large distance on the vertical axis. Felicia Hemans is the most dramatic example.

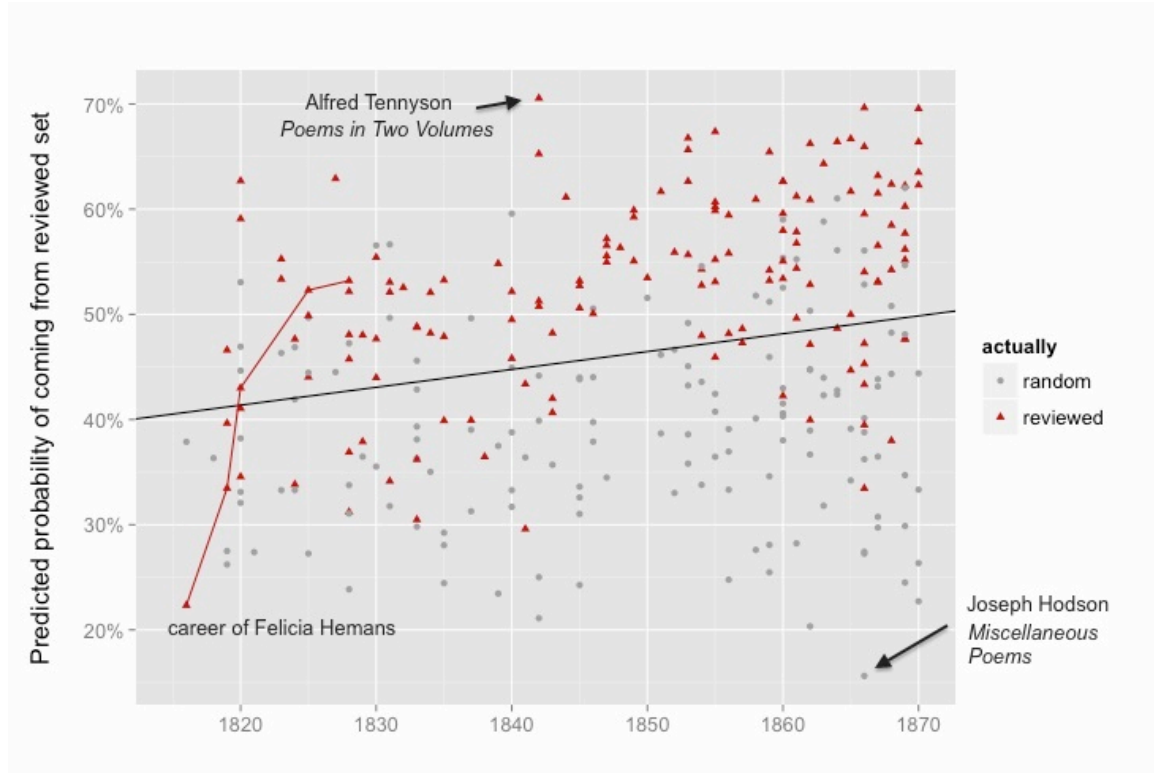


Figure 3. Felicia Hemans’ career, plotted in a space defined by reception across the whole period 1820-1919, although only volumes to 1870 are plotted. The volumes most and least likely to be reviewed are also labeled.

We have five volumes from Hemans in this dataset. The first, *The Restoration of the Works of Art to Italy* (1816, reviewed in 1820), takes a celebratory stance toward an abstract subject. “For thee, bright Genius darts his living beam, / Warm o’er thy shrines the tints of Glory stream.”³³ The poetic pleasures here are pleasures the eighteenth century might have known better how to value. The last of Hemans’s volumes, *Records of Woman* (1828), adopts a very different stance, emphasizing individual female figures at moments of crisis: “Down a broad river of the western wilds, / Piercing thick forest glooms, a light canoe / Swept with the current.” Nineteenth-century observers and contemporary scholars have generally concluded

³³ Felicia Dorothea Hemans, *The Restoration of the Works of Art to Italy: A Poem* (Oxford: W. Baxter, 1816), 12.

that *Records of Woman* was Hemans's best-received book.³⁴ Although all of her titles come from our reviewed sample, this model clearly agrees that her later books were more likely to succeed (in the century we've modeled) than the early ones. Stephen C. Behrendt interestingly argues that Hemans came under strong pressure from reviewers, around 1819, to write in a less "classical," "learned" mode, and to explore instead narratives about "the troubled regions of passion."³⁵ Behrendt believes this pressure changed Hemans's approach to poetry, but we don't insist on causality here; questions about a writer's motives can hardly be illuminated by a model of reception after her death. We've mentioned this incident merely to show what the preferences implied by our model might sound like in the mouth of a nineteenth-century reviewer. Stories about "the troubled regions of passion" are definitely what this model wants to hear.

In a sense it's scandalous that statistical models can predict poetic reception without paying attention to versification or rhyme.³⁶ The rise of *vers libre* normally plays an enormous part in our narrative of this period.³⁷ Of course, it has that central role partly because it's something critics disagreed about, so it's not clear that it would actually be useful as a predictive clue; further research will be needed to find out. The model also knows nothing about slogans like "aestheticism" or "imagism" — which are likewise central to our literary histories, because central to critical debate. We don't mean to suggest that any of these things were unimportant, but the point of this model is to give us an alternate perspective. Instead of foregrounding things that became subjects of disagreement, it foregrounds a dimension of aesthetic reception where John Keats's diction ("take into the air my quiet breath") looks more like Amy Lowell or Walt Whitman ("smoke of my own breath") than it does like Caroline De Windt's *Melzinga: A Souvenir* (1845). De Windt uses rhyme, and "the mountains /

³⁴ Paula R. Feldman, "Introduction," *Records of Woman, with Other Poems*, ed. Paula R. Feldman (Lexington: University Press of Kentucky), xi-xxix.

³⁵ Stephen C. Behrendt, "Certainly not a Female Pen': Felicia Hemans's Early Public Reception," *Felicia Hemans: Reimagining Poetry in the Nineteenth Century*, ed. Nanora Sweet and Julie Melnyk (Houndmills: Palgrave, 2001), 101, 108.

³⁶ At least, not explicitly. It's quite likely that meter does leave traces in word choice.

³⁷ For a history of meter in this period that reaches well beyond the rise of free verse, see Meredith Martin, *The Rise and Fall of Meter* (Princeton: Princeton University Press, 2012).

Frowned in [majesty sublime](#)” may ostensibly be “romantic,” but the model places Keats much closer to Whitman.³⁸ It’s true, in other words, that individual writers’ careers were shaped by overt struggles over concepts like imagism and free verse. But careers were also shaped by definitions of poetic distinction that are harder to historicize, because they changed very slowly.

Ordinarily, it’s hard for us to acknowledge this relatively stable dimension of aesthetic judgment without falling back on a universal notion of quality (or an equally dubious concept of “influence,” which might paint Whitman as somehow specifically Keatsian). One of the advantages of distant reading is that it can be more patient with historicism, revealing even slow changes as historical phenomena. The criteria of judgment revealed in this model are not universal: they are defined by patterns of reception in a particular century. When we divide that century into four parts, we get four strongly-overlapping but different sets of criteria. Even when we model the whole century as a single unit, with a single set of criteria, change is still visible — in the sense that the bar for “review-worthiness” steadily drifts upward. (Poetic language has to be more and more concrete, less and less sentimental, to actually pass the bar.) These important but gradual changes are difficult to describe persuasively with our ordinary critical toolkit.

Synchronic distinction and diachronic change

That upward drift in the model, by the way, is definitely something that needs to be explained. The pattern is durable: it occurs no matter how rigorously we balance the distribution of works across time. If we divide the century into two or four parts and train models on each, we see an upward slope within each part. Nothing about the modeling process itself compels this chronological pattern to appear. We don’t see a strongly-marked, consistent tilt if we model other social boundaries, like authorial gender. Predictions about literary prestige are skewed across time, presumably, because the difference between unreviewed and reviewed volumes is

³⁸ John Keats, “Ode to a Nightingale,” *Lamia, Isabella, the Eve of St. Agnes, and other poems* (London: Taylor and Hessey, 1820), 110. Walt Whitman, “Walt Whitman,” *Leaves of Grass* (Washington, D.C.: Smith and MacDougal, 1872), 29. Caroline De Windt, *Melzinga: A Souvenir* (New York: 1845), 16.

always analogous to the difference between works at the beginning and end of each period: so the best solution the model can find always has an upward slope.

The same pattern appears when a model of a given period makes predictions about works outside its boundaries. If we train a model on evidence from one twenty-year segment and use it to make predictions about the next twenty years, the model will reliably perceive the volumes in its “future” as more likely to be reviewed than the volumes it was trained on. We see the same pattern when we model reception in fiction. So it’s reasonable to hypothesize a general relationship between literary distinction and historical change. Diachronic change across any given period tends to recapitulate the period’s synchronic axis of distinction.

A conjecture that broad needs a few provisos. We don’t yet know with certainty that this will happen outside the period 1820-1919; we’re hypothesizing that it will. And we’re not suggesting that this is the only kind of change that can happen in literary history: many different changes are always happening, and many of them won’t be captured by a model of distinction. For that matter, there’s more than one way to model distinction. We’ve focused on the outer boundaries of literary attention, but other scholars might prefer to emphasize distinctions closer to the center of the spotlight (say, prizes)³⁹ — and those might produce a different model.⁴⁰ So it’s not as though the whole sweep of literary history has to move in any single direction. We’re just suggesting that, whenever scholars do define a linguistic proxy for social distinction in a given period, they’ll find that change *relative to that axis* moves in an upward direction during the period itself. This pattern isn’t shocking: it’s easy to imagine various reasons why it might happen. But that’s not the same thing as saying that we expected it.

It’s actually a little odd that a model trained on 1840-59 sees works from the 1870s as more likely to be reviewed than the works it was trained on. We didn’t in fact

³⁹ For the ambiguous status of nineteenth-century prizes, see James F. English, *The Economy of Prestige: Prizes, Awards, and the Circulation of Cultural Value* (Cambridge: Harvard University Press, 2005), 28-49.

⁴⁰ We actually suspect, however, that different models of distinction will turn out to be mostly congruent. Although the model we’re presenting here is technically based only on a binary contrast (reviewed, or not), hyper-canonical writers like Alfred Tennyson already do especially well within it. We also recorded reviewers’ judgments when they were plain, and those judgments do weakly but significantly correlate with this model’s predicted probabilities of review.

expect to see this, and we don't want to claim that we understand yet why it happens. We might speculate, for instance, that standards tend to drift upward because critics and authors respond directly to pressure from reviewers, or because they imitate, and slightly exaggerate, the standards already implicit in prominent examples. Synchronic standards would produce diachronic change. But causality could also work the other way: a long-term pattern of diachronic change could itself create synchronic standards if readers in each decade along the way formed their criteria of literary distinction partly by contrasting "the latest thing" to "the embarrassing past." In fact, causal arrows could run in both directions.

There are ways we could start to untangle this causal knot. It's interesting, for instance, that predicted probabilities of review correlate with authors' dates of birth more strongly than they correlate with publication dates; the sort of rapid change *within* a career we saw in Felicia Hemans's case is actually very unusual.⁴¹ But as social scientists understand all too well, causal processes are hard to trace in detail; we're not going to attempt it here.⁴²

Nor do we actually need a causal explanation of this phenomenon to see that it could have far-reaching consequences for literary history. The model we've presented here already suggests that some things we've tended to describe as rejections of tradition — modernist insistence on the concrete image, for instance — might better be explained as continuations of a long-term trend, guided by established standards.⁴³ Of course, stable long-term trends also raise the specter of Whig history. If it's true that diachronic trends parallel synchronic principles of judgment, then literary historians are confronted with material that has already, so to speak, made a teleological argument about itself. It could become tempting to draw Lamarckian

⁴¹ For a speculation to the same effect, see Moretti, "Slaughterhouse," 222. Benjamin Schmidt explains why birth dates might matter in "Age Cohort and Vocabulary Use," *Sapping Attention*, April 11, 2011. <http://sappingattention.blogspot.com/2011/04/age-cohort-and-vocabulary-use.html>

⁴² Judea Pearl, *Causality: Models, Reasoning, and Inference*, 2nd edition (Cambridge: Cambridge University Press, 2009).

⁴³ "Modernist poetry, likewise, focused upon the concreteness of images and objects, as poets responded to a diminished romanticism." (This belief is widely shared, but our results suggest it needs to be reversed by replacing "responded to a diminished" with "extrapolated.") Robert P. McFarland, "Introduction," *Film and Literary Modernism*, ed. Robert P. McFarland (Newcastle upon Tyne: Cambridge Scholars Publishing, 2013), 8.

inferences — as if Keats’s sensuous precision and disillusionment had been trying to become Swinburne all along.

We hope readers will remain wary of metaphors that present historically contingent standards as an impersonal process of adaptation. We don’t see any evidence yet for analogies to either Darwin or Lamarck, and we’ve insisted on the difficulty of tracing causality exactly to forestall those analogies.⁴⁴ On the other hand, literary history is not a blank canvas that acquires historical self-consciousness only when retrospective observers touch a brush to it. It’s already full of historical observers. Writing and reviewing are evaluative activities already informed by ideas about “where we’ve been” and “where we ought to be headed.” If individual writers are already historical agents, then perhaps the system of interaction between writers, readers, and reviewers also tends to establish a resonance between (implicit, collective) evaluative opinions and directions of change. If that turns out to be true, we would still be free to reject a Whiggish interpretation, by refusing to endorse the standards that happen to have guided a trend. We may even be able to use predictive models to show how the actual path of literary history swerved away from a straight line. (It’s possible to extrapolate a model of nineteenth-century reception into the twentieth, for instance, and then describe how actual twentieth-century reception diverged from those predictions.) But we can’t strike a blow against Whig history simply by averting our eyes from continuity. The evidence we’re seeing here suggests that literary-historical trends do turn out to be relatively coherent over long timelines.

Potential problems with the model

This project is not a history of literary reviewing itself; it uses reviewing mainly to create social contrasts that illuminate broader questions about the pace and direction of literary change. So there are many aspects of the social history of reception we won’t try to cover. A model is a simplified representation of the world: the whole point is to leave some things out. Problems arise, however, when a social variable is not simply left out of a model, but used in unacknowledged ways to shape

⁴⁴ Here we part company from Franco Moretti, *Graphs, Maps, Trees* (London: Verso, 2005).

the model's conclusions. For instance, if poems by women rarely got reviewed, and women also disproportionately used a particular vocabulary — say, a language of sentiment we've seen the model reject (“homes”, first-person plural, etc.) — then our model might be confounding literary prestige with gender. Its predictions about prestige would seem accurate only because it was leaning on the depressingly reliable assumption that works by women won't get reviewed.

In this dataset, divisions of gender and nationality are large enough that they could easily produce that kind of confounding effect, so we've recorded social information about the author of each volume to check for interactions. In the case of gender, we've checked carefully and can say confidently that we're not seeing large distortions. Women are underrepresented in this dataset, contributing only about a quarter of the works overall.⁴⁵ But they are distributed roughly equally across the reviewed and random samples — with a slightly (but not significantly) stronger presence on the reviewed side of the boundary. The model's predictions for women are just as accurate as those for men, and if you run the modeling process on a dataset restricted to women, it works just as well (or is in fact slightly more accurate). The weights the model assigns to specific words do of course change if we use only evidence from women writers, but the patterns remain broadly the same. Abstract ideals, including “home,” still reduce the likelihood of review. Concrete, troubling images still help. (On the other hand, feminine pronouns may become slightly less significant as a positive force; they seem, perversely, to help men more than women.) In short, gender certainly changed the boundaries of poetic distinction in ways worth studying, but we see no evidence that it undermines the broad conclusions drawn in this article.

The question of nationality is more vexed, because we're drawing our “random” authors from HathiTrust, which mainly aggregates the collections of large American libraries. As a result, we tend to have more obscure volumes from San

⁴⁵ Women may have been even more severely underrepresented in the ranks of poetry reviewers. See Joanne Shattock, “Reviewing,” *A Companion to Victorian Poetry*, eds. Richard Cronin, Alison Chapman, and Antony H. Harrison (Oxford: Wiley-Blackwell, 2007), 381. See also Nicola Diana Thompson, *Reviewing Sex: Gender and the Reception of Victorian Novels* (New York: New York University Press, 1996).

Francisco or Cincinnati than we do from, say, Manchester. American authors are overrepresented in the random sample, and their works are probably more obscure on the whole than randomly-selected works by British writers. The upshot of this is that the model makes more accurate predictions for Americans; when writers of all nationalities are mixed in a single model, American and Canadian writers are correctly placed 83.5% of the time, but writers from the United Kingdom and Ireland, only 74.7% of the time. So nationality probably is a confounding factor in this model, although it doesn't by any means explain away all the effects we're observing. As we expand this project into the twentieth century, we may address the problem by focusing on American authors and periodicals.

What became of our original hypothesis?

The original goal of this experiment was to test whether reviewed and random samples would become easier to differentiate as time passed. Critical tradition suggested that distinctions between popular and elite poetic culture had hardened “over the course of the nineteenth century, as the increasingly centralized media and entertainment industries interacted with the growth of education.”⁴⁶ So we didn't necessarily expect to see a systematic differentiation of poetic styles before 1850. We hoped that the gradual emergence of that sorting principle would give us a way to trace the separation of elite literary culture from the rest of the literary field.

The imprecision of “the rest of the literary field” is deliberate. Our random sample is not guaranteed to capture anything as specific as “popular culture” or “mass culture.” It does contain some collections of popular newspaper poetry, like *Poems for the Million* (1871). And it has quite a number of odd things you might call outsider art — for instance, *Po Crosby's Labor Odes; Containing Twenty-Five Beautiful Odes, Composed in the Interest of the Toiling Masses* (1889), printed in East Saginaw, Michigan, and sold by the author for ten cents. But there are also many authors in the random sample who moved in genteel social circles in Philadelphia or London, and just never happened to

⁴⁶ Janet Gray, “Popular Poetry”, *Encyclopedia of American Poetry: The Nineteenth Century*, ed. Eric Haralson (New York: Routledge, 2001), 347.

become particularly distinguished as poets. Labels like “popular” or “middlebrow” would be too specific for this sample; it combines many different things to create a ground for the figure of distinction defined by selective periodicals.

Our original plan was to compare twenty-year periods: in the period 1840-59, we guessed a model of literary distinction might only be 60% accurate, but as we proceeded toward the twentieth century it would presumably get more and more accurate, as the styles aimed at different reading audiences became clearly differentiated. We hoped for 80% accuracy by the twentieth century — which is exactly what we found there. But what we didn’t find was a significant blurring of boundaries in earlier periods. There is only a slight change. Our data divides into two equally-sized parts in 1876. A model limited to volumes published before that year is 77.3% accurate; one limited to volumes from 1876 forward is 80.5% accurate. So maybe the standards of literary distinction did become more clearly marked in the late nineteenth century — or maybe the later period just contains more deeply-obscure American books. Since American volumes are slightly easier to sort, an increase in the proportion of American books could make the increase in differentiation seem slightly larger than it really was.

On the other hand, we can’t find many sources of error that would have made this change seem smaller than it was. And we certainly came into this project expecting to see more change. We had been told to expect “an increasing tendency throughout the nineteenth century for poetry to become a discourse of distinctly high culture, a class discourse.”⁴⁷ Instead we see an elite poetic culture that is already strongly differentiated from other forms of poetic production by the second quarter of the nineteenth century. Prominent periodicals had already reached a stable consensus about the sort of poetic language worth reviewing, and we can already identify poetic discourses (call them popular, sentimental, or hortatory) that are excluded by that consensus. But if that’s true, why have literary historians agreed so widely that the

⁴⁷ David G. Riede, “The Victorian Era,” *Columbia History of British Poetry*, ed. Carl Woodring, James S. Shapiro (New York: Columbia University Press, 1994), 445.

separation of “elite” and “mass” culture at the end of the nineteenth century was a new thing?

Hundreds of shrewd observers can’t be entirely wrong. We suspect received wisdom about this “great divide” has been misleading only when it forgets to specify which side of the divide was transformed more deeply. As book historians have recognized, the new literary institutions that emerged in the second half of the nineteenth century were mostly located at the popular end of the market — dime novels, yellowbacks, pulp magazines, and even new kinds of popular verse (e.g., John Timberman Newcomb’s “Invasion of the Tinsel Rhymesters.”)⁴⁸ These forms of differentiation weren’t targeted in the present study, which focused instead on the contrast between elite literary taste and “everything else” (a contrast that, it turns out, was strong all along). When we address fiction, we plan to use collections of bestsellers and pulp fiction, in order to explore differentiations within the very broad part of the literary field that is reduced here to a single random sample.⁴⁹

Still, it’s worth knowing that the standards of elite poetic taste had already consolidated by the second quarter of the nineteenth century, and that they remained largely stable through 1919. Readers may wonder how our current pedagogical canon would fare, judged by the same standards. That’s a question we can also answer. For instance, we can create a set of poets whose works are included in the most recent *Norton Anthologies of English (and American) Literature* — adding volumes to our sample where now-canonical authors were missing, and using estimated dates of composition for writers like Emily Dickinson who went unpublished in their own lifetimes.

⁴⁸ John Timberman Newcomb, *Would Poetry Disappear? American Verse and the Crisis of Modernity* (Columbus: Ohio State University Press, 2004), 47-103.

⁴⁹ It would be also interesting to combine this model with approaches that give a more detailed account of the relationships between literary institutions. In particular, it might dovetail nicely with a history of publishing. See Natalie Houston, “Toward a Computational Analysis of Victorian Poetics,” *Victorian Studies* 56.3 (2014): 498-510. Richard Jean So and Hoyt Long, “Network Analysis and the Sociology of Modernism,” *boundary 2* 40.2(2013): 147-82.

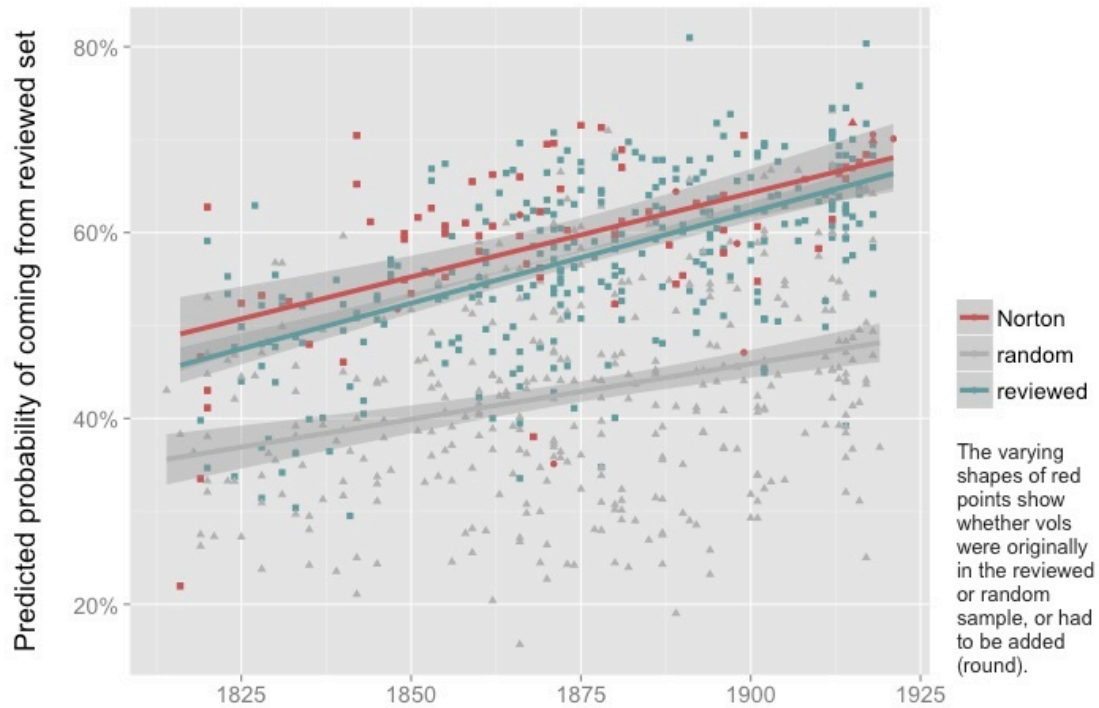


Figure 4. Anthologized writers, compared to two other samples, in a model of reception 1820-1919.

The lines plotted here are no longer lines of division between categories; they are the central trend lines for each category, plus a shaded 95% confidence interval (indicating 95% confidence that the true trend line falls somewhere in the shaded area). The separation between “reviewed” and “random” categories is very clear — and there may be a hint that those categories are diverging more and more over time. But the trend lines for volumes selected by 19th-century reviewers, and by editors of 21st-century anthologies, are not separable. This doesn’t imply that our pedagogical canon is in every way a typical subset of poetry reviewed in 19th-century periodicals. It tells us only that the model finds these samples indistinguishable in terms of the particular metric plotted here: their degree of similarity to reviewed or randomly-selected volumes. Canonical authors who weren’t reviewed in their own lifetime, like Emily Dickinson, don’t change this pattern as much as we might expect: the model finds that Dickinson’s poems, for instance, actually align very well with her century’s

implicit poetic standards, and it assigns them a well-above-average chance of being reviewed — had she published them.

We probably shouldn't take the model's counterfactual opinions about particular authors very seriously, but Figure 4 does demonstrate a broader point: the standards of poetic distinction we've modeled in this study were remarkably durable, not only from 1820 to 1919, but because they continue to guide academic attention today. Recovery efforts have not done much yet to overturn the devaluation of styles that were excluded from literary reviews in the nineteenth century.

Conclusion

A lot of descriptive work remains to be done in literary history, because we still know fairly little about macroscopic patterns in this field. Literary historians have often generalized about the pace of change, for instance, contrasting epochs of relative stability to the “revolutions” that separated them.⁵⁰ But do we have evidence for those claims? It's hard to describe the pace of literary change over long periods. It's easier to point to manifestoes, and that may well be how Romanticism and Modernism got their reputation as moments of radical discontinuity. With ordinary historical methods, it's even harder to describe changes in standards of reception. If you're looking at works on one side of a boundary, the boundary itself is invisible.

We've suggested that contrastive sampling can give scholars a way of describing the boundary between works consecrated by publicity and those that remained obscure. At least in the case of Anglo-American poetry, the implicit standards governing reception may have remained relatively stable between 1820 and 1919. Admittedly, we haven't tried to model versification, and versification may have been the most volatile dimension of poetic practice. But we also didn't need the evidence of versification to make broadly accurate predictions about reception. We're not suggesting, after all, that poetry remained unchanged across this period: there were certainly many changes not reflected in this model. But the nature of the difference between prestigious and obscure works remained largely the same. You can

⁵⁰ See the works by Fallis and Greenblatt, et. al., cited above.

model poetic distinction across this whole century, in fact, using a single set of evaluative criteria that are applied more exactly as time passes.

We haven't completely described those criteria. But a few broad outlines are clear: abstract ideals and celebratory sentiments were consistently devalued, in favor of concrete (and often dark or "troubled") imagery, from 1820 to 1919. Though we haven't had space to discuss fiction in this article, we can't resist remarking — for whatever it's worth — that the patterns we're seeing in that genre so far are similar. We haven't explained the origin of these patterns yet, and hardly have space in the pages that remain to defend an explanation. But literary scholars may not actually find it hard to explain macroscopic patterns once the patterns themselves are sketched at full length. We already have many explanatory theories about literary history, after all; if a pattern sprawls across a century it's almost certainly going to be entangled with some of them. What we haven't had until recently is a good way of distinguishing trees from forests — separating debates that happened to make noise in particular decades from durable structural patterns.

If we need to explain evaluative standards in a way that can embrace a whole century of literary history at once, the range of plausible explanations narrows a great deal. One strong contender, at least, may be Pierre Bourdieu's notion that the values of the nineteenth-century literary field were assertions of the field's own social autonomy. As we suggested a few years ago, Bourdieu's theory accounts well for literary elites' insistence on concrete diction.⁵¹ In the world at large, the prestige and economic value of literacy are associated with abstract (for instance, legal) vocabulary. So, as the literary field makes a bid for full autonomy, that value has to be reversed: only an insistently plain language of sense experience can count as eloquence here.⁵² Bourdieu's notion that nineteenth-century aesthetic values were defined through

⁵¹ Ted Underwood and Jordan Sellers, "The Emergence of Literary Diction," *Journal of Digital Humanities* 1.2 (2012). <http://journalofdigitalhumanities.org/1-2/the-emergence-of-literary-diction-by-ted-underwood-and-jordan-sellers/>

⁵² The significance of this reversal tends to be obscured, because poets from William Wordsworth to T. S. Eliot have been eager to present themselves as rejecting conventional "poetic diction" and embracing the "ordinary language of men." But the special notion of ordinariness they embrace never actually constitutes a move toward the linguistic practice of non-poetic genres. (In fact, when we compare this whole dataset of poetry to nonfiction, we find that distinction within the poetic field correlates with distance from nonfiction.)

rejection of heteronomy might also make sense of our model's most perplexing feature: its rejection of all positive ideals, and preference for things that are (for instance) "dull," "dumb," "harsh," or "dead." In fact, Bourdieu's discussion of Baudelaire already shows how conventions of negativity could have sprung from self-conscious rejection of a dominated position.⁵³ All we've done here is to suggest that Baudelaire's paradoxical aesthetic of degradation was not as idiosyncratic as it looks. It was also diffused across a century and embraced tacitly as a definition of poetic prestige in Victorian periodicals whose editors perhaps rarely took laudanum.

But as we've already conceded, the task of inventing explanations is not the hard part of distant reading. Once we get a clear view of macroscopic patterns, it may not be difficult to see how they fit some of the dozens of conflicting theses scholars have already proposed. It may even become clear that apparently conflicting trends actually form interlocking parts of a puzzle: the chilly reticence of modernism, say, may snap together with the histrionic Victorian nostalgia it was supposedly reacting against, once we see how both things can be explained by a single linguistic model. The challenge at this scale is not to come up with new theories, but to show how they fit together.⁵⁴ It's hard to hold a century of writing in your mind as a single gestalt; even if you could do that, it would be hard to persuasively communicate your vision, because a rhetoric organized by individual examples isn't good at distinguishing continuity from change across a hundred years. We've tried to show that probabilistic predictive modeling can help literary historians come to grips with these questions. We've focused on describing continuity, but the same techniques could be useful for describing change — since we can, for instance, extrapolate a model into its own future, and show where it fails. Predictive models of literary prestige can also

⁵³ Pierre Bourdieu, *The Rules of Art: Genesis and Structure of the Literary Field*, trans. Susan Emanuel (Stanford: Stanford UP, 1995), 60-68, 77-81. This is still speculative. It's also likely, as Laura Mandell pointed out in discussion, that melancholy has been poetically prestigious since (at least!) the early modern era. We might need more nuanced analysis to distinguish generalized melancholy from the anti-poetic poetics Bourdieu associates with Baudelaire. On this topic, see also Andrew Goldstone, *Fictions of Autonomy: Modernism from Wilde to de Man* (Oxford: Oxford University Press, 2013), 174-77.

⁵⁴ In general, it doesn't make sense to characterize algorithmic models as "surface readings"; the whole point is to draw inferences about connections that aren't open to inspection. But we do agree that the challenges of historical description have been underestimated. Stephen Best and Sharon Marcus, "Surface Reading: An Introduction," *Representations* 108.1 (2009): 1-21.

contribute to historical explanation by revealing a relationship between synchronic and diachronic dimensions of history, allowing us to say which aspects of change really paralleled, or diverged from, prevailing standards.

We don't think these methods need to be at odds with more traditional kinds of interpretation. We can model a social boundary, for instance, but also interpret it by looking closely at the literary pleasures that flourished on either side. (The proliferation of case studies may be limited by readerly patience, but it isn't limited by quantitative method as such.) The methods used here are also open to our discipline's familiar modes of theoretical debate, because statistical modeling is an explicitly hermeneutic enterprise. Writers who dismiss models as "merely empirical" or "positivist" strategies are announcing mainly that they haven't yet engaged the debate about the interpretive limits of modeling that flourishes in other disciplines, and increasingly our own.⁵⁵

Distant reading conflicts with literary historians' existing goals only in the practical sense that it does take a lot of time. This article only discusses a medium-sized corpus of 720 volumes, but to create a contrastive touchstone that was meaningfully different from reviewed volumes, we had to sample a much larger collection. Like other researchers in this field, we share our code and data, and we hope over time that practice will help reduce bibliographic and technical obstacles to a project like this.⁵⁶ But distant reading is going to continue to take time; figuring out what it really can or cannot do will require patience.⁵⁷

University of Illinois, Urbana-Champaign, May 2015

⁵⁵ One central question is how opaque, or interpretable, the models produced by machine learning actually are. Compare Breiman, "Statistical Modeling: The Two Cultures" to Galit Shmueli, "To Explain or to Predict?" *Statistical Science* 25.3 (2010): 289-310. See also Alan Liu, "The Meaning of the Digital Humanities," *PMLA* 128 (2013): 409-23, and Andrew Piper, "Novel Devotions: Conversational Reading, Computational Modeling, and the Modern Novel," *New Literary History* forthcoming 2015. Preprint: http://piperlab.mcgill.ca/pdfs/Piper_NovelConversions_Preprint.pdf

⁵⁶ See Underwood, "Page-Level Genre Metadata."

⁵⁷ The collections of data and metadata that underpin this argument were made possible through grant support from the National Endowment for the Humanities, the American Council of Learned Societies, and the Social Sciences and Humanities Research Council. Any views, findings, conclusions, or recommendations in this publication do not necessarily reflect those of the funding agencies.

METHODOLOGY

The stylistic protocols of literary criticism don't encourage a detailed description of data preparation and statistical modeling, so we've gathered a lot of methodological caveats here.

Selection of periodical venues

Our initial choice of periodicals was guided by the instincts of an informal advisory board, but how can we know that they correctly identified the most prestigious venues? We can't — but we do have ways to know whether any particular venue is well-aligned with the larger group. A statistical model allows us to compare the *Edinburgh Review* to *Tait's Magazine*, and say confidently that the works reviewed by the *Edinburgh* were on the whole *more like* the kinds of works that would be reviewed by other leading periodicals throughout the century.

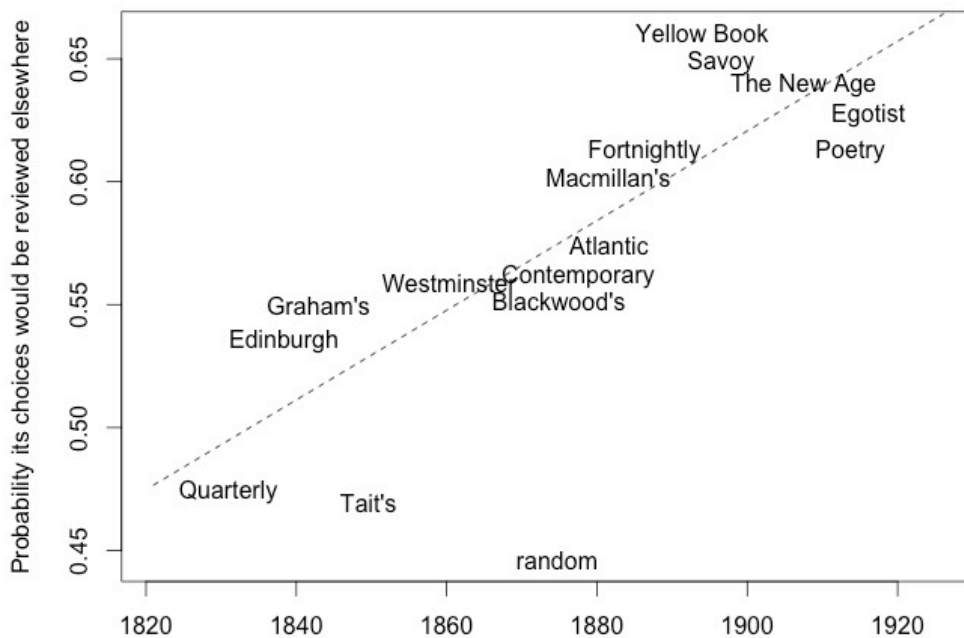


Figure 5. Estimating levels of agreement between periodicals.

To produce Figure 5, we averaged the predicted probabilities of review for all the works we recorded as reviewed in a given periodical. This gives us, roughly, an estimate of how well its standards match the standards of the model as a whole. Of course, this is systematically unfair to early publications, since the model always (for some unknown reason) sees early works as less likely to be reviewed. So we've also averaged all the dates for works reviewed in each periodical, and plotted the publication name at that mean date. Then we do a linear regression for all the publications (not counting the "random" works at the bottom). This gives us a rough sense of which publications are above or below the line: better-aligned with the model (or less so) than would be expected for their era. But in truth the differences between most periodicals are not huge, with one or two exceptions.

In particular, we used Figure 5 to confirm a suspicion that *Tait's Edinburgh Magazine* was unlike the other periodicals in our project. We had relied on *Tait's* fairly heavily, since it was difficult to find other sources for reviews in the period 1840-1859. But that worked because *Tait's* reviews a relatively large number of titles — perhaps a little indiscriminately. (It's a monthly, in an era when the most prestigious reviews were quarterlies.) Our reading of *Tait's* had already led us to suspect that might be a problem. Looking at quantitative evidence that its standards were not well-aligned with other periodicals in our sample, we decided to exclude all reviews from *Tait's*. (After all, our goal was not to collect a random sample of different magazines: it was explicitly to find a group of periodicals that represent the tastes of a coherent "literary elite," if such a thing could be constructed.) We made up for the loss of *Tait's* by adding *Graham's Magazine* and a larger sample of reviews from the *Westminster*. This change — plus more accurate dating of volumes — lifted the overall accuracy of our model from 74.4% to 79.2%, so choices about particular venues do clearly make a difference. But it's a debatable choice, which is why we've explained it at length here. In particular, taking out *Tait's* made it difficult to achieve our original plan of distributing volumes evenly across the timeline; coverage now gets less dense as you move backward in time.

Sampling and data preparation

All of the texts we used were drawn from HathiTrust Digital Library and processed in the same way, so that there would be no mechanical differences between the reviewed and random samples — only differences of social provenance. Spelling was normalized to modern British practice. Volumes of poetry often include prose introductions, or front and back matter; this was trimmed using publicly-available metadata.⁵⁸ That process has a known level of error, but there's no reason why errors should be distributed differently across the samples.

The “reviewed” sample used in this article was created mostly by Jordan Sellers. He paged through periodicals looking for reviews of poetry and fiction published relatively close to the review date, and of which we possessed a digital text. Since we have digital copies of about 53,200 volumes of poetry in the period, the limitation was not unduly constraining. But if we didn't have a copy of a particular title or edition, we often made do with another volume by the same author. Since publication could also be a sign of editorial taste, we recorded a small number of works (about 5% of the total sample) that were published rather than reviewed in selective venues. This was a debatable choice, since it mixes different forms of editorial judgment. But it had in any case a small effect.

Ted Underwood created the random sample. If we wanted to create a list of absolutely unreviewed works, we would have had to start by making an exhaustive list of everything ever reviewed in particular periodicals, in order to exclude those titles. But we were satisfied to call this instead a random sample. It's not necessarily a list of works that were *never* reviewed, just a list of things that are likely to have been reviewed less often in our sample of venues. In fact, when a stray volume from the random set turns up near the top of our model's list of books likely to be reviewed, it does turn out that many of those authors are reasonably well-known (Rupert Brooke, Elaine Goodale Eastman). You could say that they wound up in the random sample

⁵⁸ Underwood, “Page-Level Genre Metadata.”

“by mistake,” but this is an expected mistake, and a relatively benign one, since it would only tend to understate the actual strength of the patterns described here.

Although we selected these volumes literally at random, we did exclude authors who were already in our reviewed sample for a given genre. (We also excluded most anthologies by multiple authors.) The model is thus fundamentally distinguishing authors rather than individual works; you might say it distinguishes authors who we know became prominent at some point in their lives from a sample of those who perhaps never did. For each volume we have three or four dates: a date this volume was actually published, a date the title was *first* published, for reviewed volumes a date of review, and sometimes the year of the author’s birth. (The latter three have to be constructed manually, and it’s not always possible to find a birth date for obscure authors.) We’ve looked at the data from all of these angles; the dates used here are first dates of title-publication, but plotting the volumes by author’s date of birth can also be revealing. It’s difficult to *select* titles randomly by date of first publication (which is not recorded in Hathi metadata), so we had to create a random sample larger than 360 volumes, manually record dates of first publication for them all, and then winnow the list by letting our script align “reviewed” and “random” as closely as possible.

Training and interpreting models

To create the models described here we used regularized logistic regression, a well-known learning algorithm that has the advantage of being relatively interpretable. But “relatively” is the key word in that sentence: we don’t claim to have fully explained this model, and there are lively debates between statisticians and computer scientists about the question of interpretability. Logistic regression doesn’t give interpreters a huge number of different knobs to tune, which is probably a good thing. But the regularization constant is a significant detail, because it can make trends over time seem more or less dramatic; we’ve limited our interpretive flexibility here by choosing a constant, and a number of features, that maximized predictive accuracy on the data.

A regression model with 3200 variables is not guaranteed to be transparent. The coefficient assigned to a word tells you only how variation in that word's frequency will affect a prediction. It's not necessarily a measure of statistical significance, because variables can interact in odd ways. A group of strongly predictive words that always appeared together could end up with small coefficients because predictive weight got "shared" across the group. For this reason, among others, it's risky to place a lot of interpretive emphasis on single words that happen to be near the top or bottom of a list; instead, we've tried to emphasize broad patterns. But it would also be an exaggeration to portray this model as a mysterious black box. In practice the model coefficients line up very well with a simple question like "which words are most strongly overrepresented in reviewed volumes?" (Spearman's correlation between model coefficients and a signed version of Dunning's log-likelihood comparing the two samples comes to $r = 0.887$.) In fact, the broad patterns that characterize reviewed volumes (concrete language, darkness of tone) could also have been identified using direct corpus comparison. The point of predictive modeling is that it allows us to measure the strength of a boundary, and say whether that boundary was changing over time.

In training the model we "normalize" word frequencies by the standard deviation for each word (across the whole dataset). So when we use the model to illuminate specific passages, we also divide coefficients by the standard deviation. This tells us, roughly, how much a *single* occurrence of a given word would affect the model's prediction, which is what we're trying to dramatize when we quote a passage. We've rendered words in red if they're in the top 1300 features by this metric, and colored them blue if they're in the bottom 1300. The main weakness of this strategy is that it understates the aggregate importance of common words. We mentioned that feminine pronouns contribute to a poem's odds of being reviewed, and that first-person plural pronouns detract. But there are a number of other syntactic preferences latent in the model. A paratactic style is prestigious ("and," "but," "or"). The future tense is not. These rhetorical patterns are harder to interpret than the thematic patterns we've foregrounded, but they could be at least as important.

BIBLIOGRAPHIC DATA

Bibliographic information about the volumes we used is available on Github at github.com/tedunderwood/paceofchange. That site also includes our model's predictions for all volumes, and the weights it assigned to different words. We've also shared our code and raw word-frequency data for the volumes, so readers who have Python 3 (with the scikit-learn module) can replicate our results.

Readers might also be interested to see what the two halves of the dataset look like when modeled separately. Note that the first of these illustrations covers more ground on the timeline than the second; so the visual impression that change is more rapid in Figure 6 might be a little deceptive.

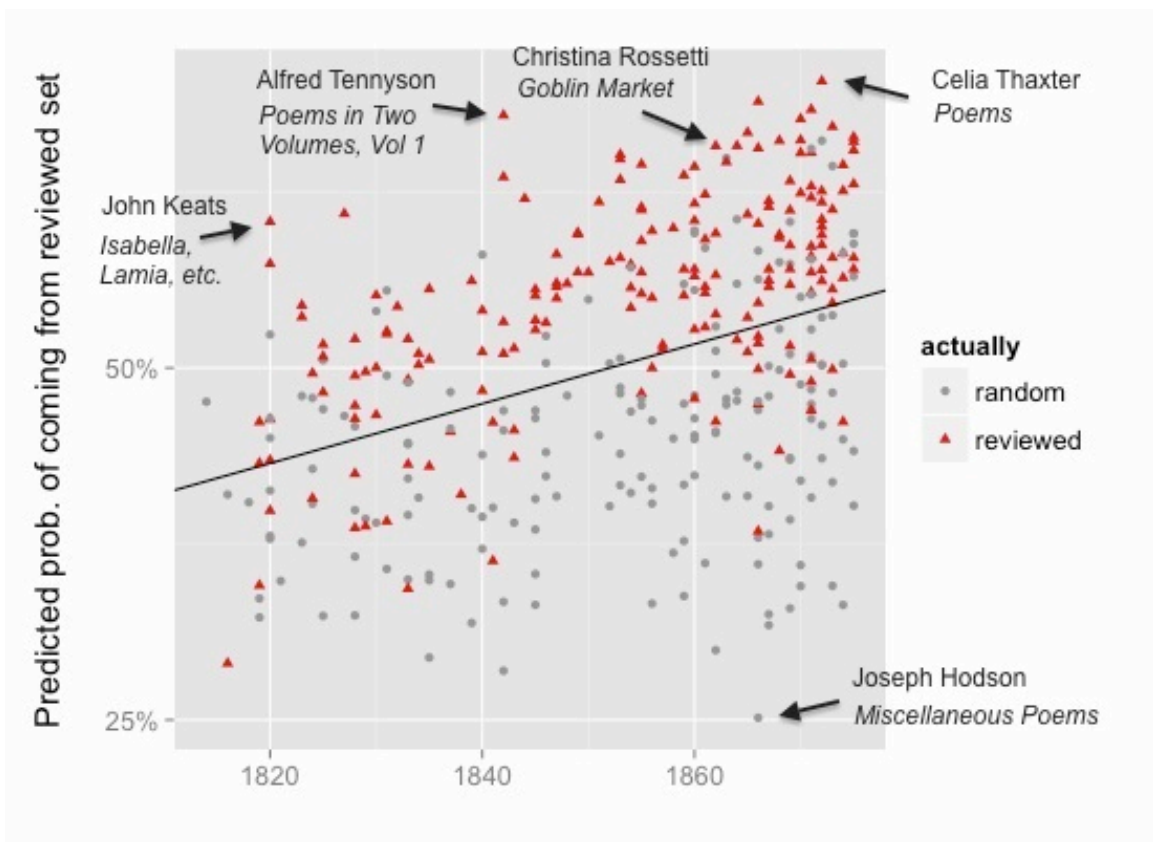


Figure 6. First half of the dataset, modeled separately.

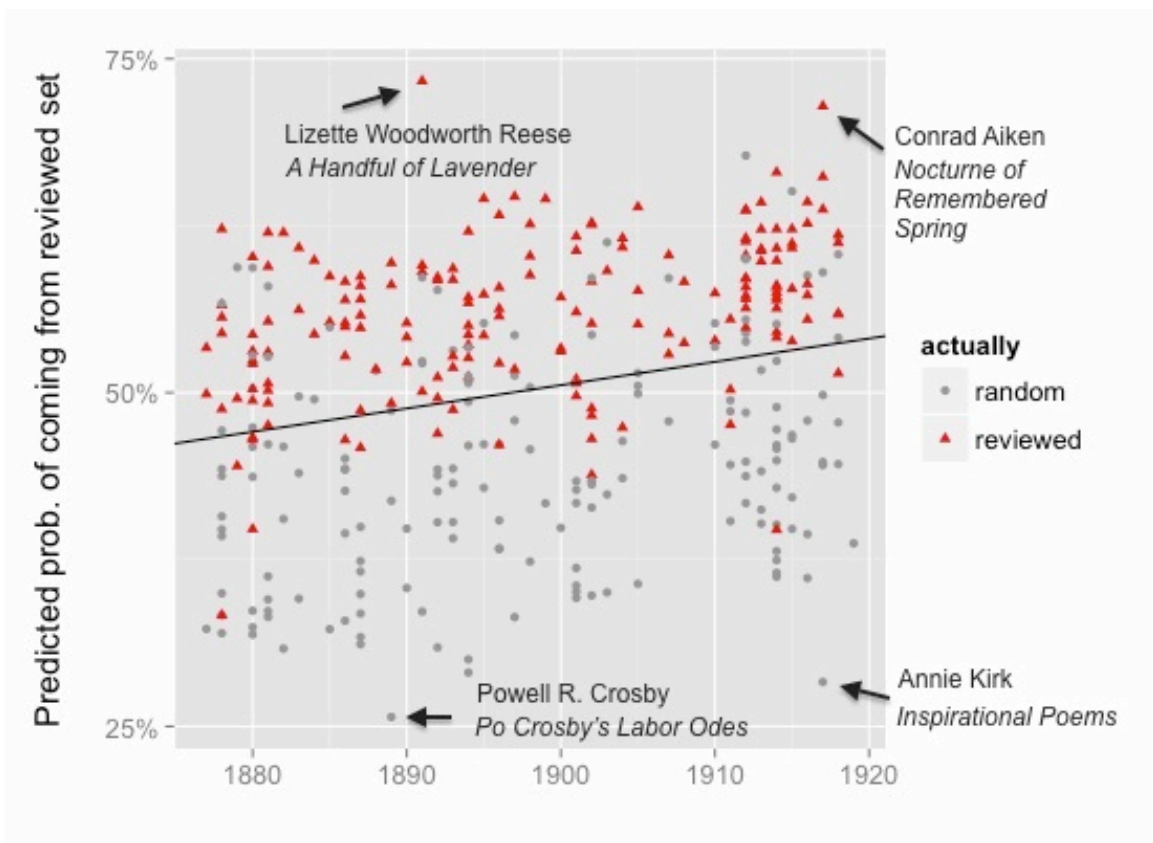


Figure 7. Second half of the dataset, modeled separately.