The *Longue Durée* of Literary Prestige

Ted Underwood and Jordan Sellers

Abstract A history of literary prestige needs to study both works that achieved distinction and the mass of volumes from which they were distinguished. To understand how those patterns of preference changed across a century, we gathered two samples of English-language poetry from the period 1820–1919: one drawn from volumes reviewed in prominent periodicals and one selected at random from a large digital library (in which the majority of authors are relatively obscure). The stylistic differences associated with literary prominence turn out to be quite stable: a statistical model trained to distinguish reviewed from random volumes in any quarter of this century can make predictions almost as accurate about the rest of the period. The "poetic revolutions" described by many histories are not visible in this model; instead, there is a steady tendency for new volumes of poetry to change by slightly exaggerating certain features that defined prestige in the recent past.

Keywords literary reviewing, distant reading, poetic diction, literary prestige, *longue durée*

Because advocates of distant reading sometimes pit the "exceptional" canon against a "mass" of unread works in a populist way, it may seem that bracketing differences of value is the whole point of their project (Moretti 2005: 3). It is true that questions about reception have often taken a backseat here, in part because distant readers have been preoccupied with trends so dramatic that the synchronic differences of prestige between works do little to change a diachronic picture.[1] But

[1] The trend that Ryan Heuser and Long Le-Khac (2012) spot in "2,958 nineteenth-century British novels," for instance, can be traced also in a more canonical sample (Algee-Hewitt et al. 2016).

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.

Modern Language Quarterly 77:3 (September 2016)
DOI 10.1215/00267929-3570634

distant reading does not by any means rule out questions of value. In "The Slaughterhouse of Literature" Franco Moretti (2000) was already posing questions about canon formation, and a recent flurry of projects show sharply increased interest in reception (e.g., Algee-Hewitt et al. 2016; DeWitt 2015).

Perceived differences of value cannot easily be studied by consulting a single list of works, however inclusive; the history of reception requires comparisons across social categories. So scholars grappling with these questions increasingly approach historical representation as a problem of contrastive sampling. Historical samples, like social-scientific ones, are always limited by a particular method of selection. No 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 (2015) have compared best sellers to novels selected by different groups of critics to discover how different forms of literary success overlapped in the twentieth century. (It turns out that they mostly didn't.) Large digital libraries can be useful for this kind of inquiry, not because library collections are appropriately balanced for all possible questions but because a large collection creates room for different sampling strategies, giving us contrastive leverage on a wide range of topics. For questions about literary production, we may want every text we can find. For other questions, we may ground our research on a subset of works defined by some measure of significance (sales, reviews, pedagogical canons). 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 is what we attempt in the pages that follow. We created two samples of poetry and fiction across the century 1820–1919: one drawn from volumes reviewed in prominent Anglo-American periodicals, the other drawn at random from the HathiTrust Digital Library, which contains at least 146,000 English-language volumes of poetry and fiction in this period, many of them relatively obscure. The contrast between the provenance of these samples allowed us to frame synchronic models of literary prestige for particular periods. But we were 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 began our inquiry with the hypothesis that a widely discussed "great divide" between elite literary culture and the rest of the literary field started to open in the late nineteenth century (Huyssen 1986: viii). We proposed to trace the emergence of that divide with predictive modeling (used in Long and So 2016). 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 diction to predict whether a literary work was reviewed in a prestigious venue, but that differentiation of styles emerged earlier than we had thought; indeed, it was stable from at least 1840. In testing this hypothesis, however, we also stumbled on broader evidence about the pace and direction of literary change. The inquiry turned out to be so productive, in fact, that this article describes only one half of it: the part concerned with poetry.

The Plan of the Experiment

Training a computer to predict whether books were reviewed in a particular set of venues was admittedly an odd strategy. We could have checked where the books were reviewed; we didn't need computers to guess for us, and we didn't really care whether computers were good at guessing. The prediction gambit was just an indirect way to answer a different question: was the social boundary between elite taste and the rest of literary production associated with any recognizable stylistic differences? The answer might be no, since we considered fourteen periodicals, presumably with different editorial standards. But if poetic prestige turned out to be associated with a distinctive style, how rapidly did that style change? Our "reviewed" poets ranged from William Wordsworth and Lord Byron to Amy Lowell and T. S. Eliot; they might not have much in common. In fact, it seemed possible that Wordsworth would resemble twentieth-century newspaper verse more closely than he resembled Eliot, in which case it might be difficult to find any model of prestige that distinguished a whole century of "reviewed" authors from a century of "randomly selected" ones. Since we expected definitions of poetic prestige to have changed rapidly, we originally planned to train several models covering twenty-year periods.

But these were hypotheses we could test as we went along. The assumption we had to make at the start was that being reviewed indicates a sort of literary distinction, even if the book is panned.[2] Scholars more commonly study reception by contrasting positive and negative reviews. That approach makes sense if you're interested in gradations of approval between well-known writers, but it leaves out many works that were rarely reviewed at all in selective venues. We believe that this blind spot matters: literary historians cannot understand the boundary of literary distinction if they look only at works on one side of the boundary.[3] So although we recorded reviewers' sentiments when they were clear, this article places more emphasis on the fact that an author was reviewed at all.

To make this strategy work, we needed to focus on periodicals that were selective about what they did review—usually quarterly, monthly, or fortnightly publications, rather than weeklies. We created an initial list of titles by quizzing friends who are scholars of this period.[4] 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 was 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.

The list of periodicals we chose appears in table 1. For each title we list the earliest and latest publication dates of volumes that we sampled from its reviews, and the number of volumes used in this study. (Although 1820 was the earliest year we sampled reviews, dates of publication can be earlier. In a few cases—about 5 percent of the volumes sampled—we used a work's publication in a periodical rather than its appearance in a review as a proxy for editorial judgment.)

[2] We don't assume that this distinction is conferred *by a reviewer.* Many editors decided what to review by looking at publishers' lists or by "puffing" their own publisher (Mason 2013).

[3] For the social theory behind the decision to model "boundaries," see Abbott 1995.

[4] We relied on advice especially from Nina Baym, Ryan Cordell, Eleanor Courtemanche, Jeff Drouin, Andrew Gaedtke, Lauren Goodlad, Matthew Hart, Deanna Kreisel, Anthony Mandal, Bruce Michelson, Justine Murison, Bethany Nowviskie, and Roger Whitson.

Table 1. Periodicals sampled

Periodical	*Publication dates*	*Number of volumes*
Atlantic	1845–1905	122
Blackwood's Magazine	1838–96	16
Contemporary Review	1877–78	2
Edinburgh Review	1819–56	36
Egoist	1912–18	17
Fortnightly Review	1863–1917	60
Graham's Magazine	1827–55	19
Macmillan's Magazine	1881	2
New Age	1907–8	2
Poetry: A Magazine of Verse	1910–16	32
Quarterly Review	1816–51	19
Savoy	1896–97	2
Westminster Review	1828–67	29
Yellow Book	1893–95	2

We also needed a sample that contained books reviewed less often. To confirm that a given book had never been reviewed in any of these publications would have been a tedious task. It was more straightforward simply to select works at random from a very large collection while excluding authors already in our reviewed sample: in practice, this turned up mostly books that were rarely reviewed. We worked with the HathiTrust Digital Library, which contains the aggregated collections of large public and university libraries: for 1820–1919, that gave us a collection of roughly 758,400 books in English, of which about 53,200 include significant amounts of poetry (Underwood 2014). A digital library is itself a sample, with selection biases. But it samples a social range much broader than the range covered by elite periodicals, and what we needed in this study was not completeness but contrast. We gathered 360 reviewed and 360 random volumes, distributed in a similar way over the timeline.

A Model of Reception

Our goal here is to assess the strength of the relationship between poetic language and reception. That is trickier than it sounds, because a pattern-finding algorithm can usually find some pattern in a finite data set.

Although a correlation between language and reception might look strong, it could be largely accidental; in effect, the algorithm has only "memorized" the quirks of particular examples. To test a model more rigorously, we have to use it to make predictions about volumes and authors it has not yet seen. If this works, we know that our model has captured a truly generalizable relationship between language and reception (Breiman 2001).

The model knows only the relative frequencies of the words each book 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 scholars, we know that words signify on multiple levels. Perhaps it won't surprise us that the choice of a single word can reflect a text's likely readership, as well as its explicit themes. In any case, computational analysis of text has relied on word frequencies because they do register many things at once.

At the grittiest mathematical level, the predictive model we create is just an equation that translates word frequencies into a probability that a particular volume came from the reviewed sample. Although we may say that it models a boundary between the samples, probability is a continuum, and a probabilistic model allows us to treat social boundaries as fuzzy gradients. We "train" the model by showing a regression algorithm 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.[5] When we show the model a new volume, by the author it hasn't yet seen, it uses the weights assigned to different words to estimate the probability that this volume was reviewed. To avoid circularity, the model never makes predictions about one of an author's books by using information about others (Sculley and Pasanek 2008). So we actually train 636 slightly different models, each one excluding books by a different author. However, since each pair of models shares more than 99 percent of their evidence, we can describe them collectively as one model of the literary field. In figure 1 we have plotted 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.

[5] We used the logistic regression functions from Pedregosa et al. 2011 and visualized results using Wickham 2009. For our own code, see Underwood 2015.

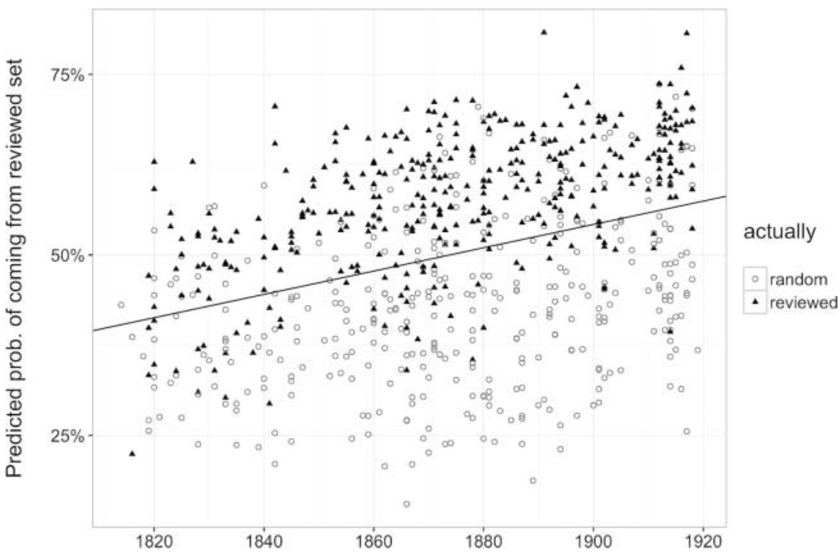

Figure 1. A model of literary prestige from 1820 to 1919

There is a lot of information here to unpack. But it is clear at a glance that our model does a reasonable job of sorting reviewed from random works—and, somewhat to our surprise, does so for the whole century at once (although it is a little less accurate before 1840). This is the first and biggest implication of the model: the verbal differences between prominent and obscure authors turn out to be stable over long spans of time. A similar result has been uncovered by the Canon/Archive project at the Stanford Literary Lab (Algee-Hewitt et al. 2016). That project uses twentieth-century critical judgment rather than nineteenth-century reviews to define prestige, but the comparisons we have made so far suggest that the criteria underlying these different kinds of prestige may be compatible (Underwood and Sellers 2015: 25–27).

But how reliable is this model? Normally, we would evaluate a model of this kind using the 50 percent line in the middle of the *y*-axis; the model predicts that everything above that divide probably came from our "reviewed" sample. Evaluated in that simple way, the model would perform moderately well, with 77.5 percent accuracy. But we can get better results by acknowledging the odd fact that the whole collection drifts upward as historical time passes. If we consider publication date as a factor and use the slanted black line to divide the data set, the model is 79.2 percent accurate.[6] Technically, the upward drift 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, since the upward drift suggests that historical change across this century moved in the direction of the standards that govern reception for the century as a whole.

Before drawing any inferences from that detail, we should more fully assimilate the fact that reviewed and random volumes can be separated with 79.2 percent accuracy. How good is that, and how good would it have to be before we called it meaningful?

First, why would anyone expect this sort of prediction to work at all? Algorithms can often infer genre with greater than 90 percent accuracy, which is why we rely on spam filters to detect advertisements in our e-mail. But in this case we are making inferences about an event that would have happened to a text only *after* it was written. Decisions about reviewing could have been made by scores of people on opposite sides of the Atlantic, guided by different standards of poetic quality (or perhaps by other factors—politics, notoriety, personal favors).

The point of trying to infer reception from poetic diction is not to automate a task that human beings find easy, like filtering spam from e-mail, but to discover whether diction is meaningfully related to reception at all. The answer to that question could be interesting, even if it turned out to be no. (Accuracy close to 50 percent would amount to a no, because in an evenly divided data set that could be random.) A model of reception would also be interesting even if it turned out only to be stable across short periods, as we initially expected.

Scholars propose different dates, but almost everyone agrees that poetic standards changed dramatically in the nineteenth century. W. B. Yeats dated the "revolt against Victorianism" and against "the poetical diction of everybody" to the 1890s (Fallis 1976: 89). Americanists sometimes identify the beginning of the change with Walt Whitman; the editors of *The Norton Anthology of English Literature* locate a "poetic revolution"

[6] We created "the slanted black line" very simply by running linear regression on the whole data set to identify a central trend.

in "the years leading up to World War I" (Greenblatt et al. 2006: 1834). It appeared in any case likely that poetic diction would be different in 1915 than it had been in 1850, and possible even that prestige would be generated in 1915 precisely by avoiding the diction that had counted as prestigious in 1850. Because we began by assuming that distinction was created by this sort of moving boundary between avant- and arrière-gardes, we expected that we would need to train different models to capture the logic of poetic distinction in different periods.

In practice, however, we found that we could separate these samples most accurately by treating the whole century as a single unit, organized by a single set of standards.[7] In other words, you can use the same list of prestigious or banal words to distinguish Byron's *Prophecy of Dante* from less prominent works in 1821 and to distinguish 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 about the outer boundary of distinction: the question of what kind of writing is even worthy of notice.

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 our random sample. The volumes of poetry we don't usually read must be united by some obvious feature. Maybe they're all religious? Or 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

[7] 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 all models are based on the same number of volumes.

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 percent of the time. It is 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 is no difference between the samples screamingly obvious to human readers.

So how can a statistical model be right 79.2 percent of the time? And what *was* the secret to getting reviewed in this period? We have to content ourselves with a rough answer. The model we have trained uses thirty-two hundred variables—the frequencies of the thirty-two hundred 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 is not required to map onto any single idea. Moreover, we may not be able to characterize the effect of any single word with great precision, since variables interact with each other in tricky ways.[8] A list of the top ten words that individually have the largest effect on the model's predictions might not tell us very much.

But if we are willing to back up and look instead at broad patterns, it is possible to grasp the general logic of a model like this by reading a few revealing passages. To bring out the big picture, we have roughly divided the model's variables into three groups: the top thirteen hundred words, which markedly increase a poem's perceived likelihood of being reviewed, are represented in boldface. The bottom thirteen hundred, which markedly decrease that likelihood, are italicized. All others are

[8] The tension between explanation and prediction involves technical details ("multicollinearity"), but it is also a large and interesting philosophical question. Compare Breiman 2001 to Shmuéli 2010. See also Goldstone 2016, which comments on a working draft of the present article, rightly stressing that the specific ordering of features in a model can be volatile, and that predictive models are not causal explanations. (When a particular word increases the probability that a text was reviewed in this model, we're describing an association, not claiming that it literally convinced anyone to write a review.) On the other hand, cross validation on held-out evidence makes predictive models more robust than the mere correlations readers may be accustomed to discounting: these are generalizations whose portable value has been tested.

typeset normally (this includes words too rare to be included in the model, as well as the middle six hundred words, which don't individually have a huge effect). We start with the conclusion of Christina Rossetti's (1865) "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.

One can detect traces of the advice that writing teachers still give: "Use definite, specific, concrete language" (Strunk and White 2007: 37). 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 is 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 kind of writing that gets reviewed.

Even in that capacity the model is 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 might as well 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 is rightly confident will never be reviewed. Many of these have some inspirational or hortatory purpose; they are about equally divided between religious and political topics but share a reliance on positive emblems of collective emotion. In *Memorial or Decoration Day*, for instance, George Loomis (1891) invokes "those who battled for these *homes* of *ours*, / And *precious* blood on *Freedom's altar* shed." By no means all the volumes in 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

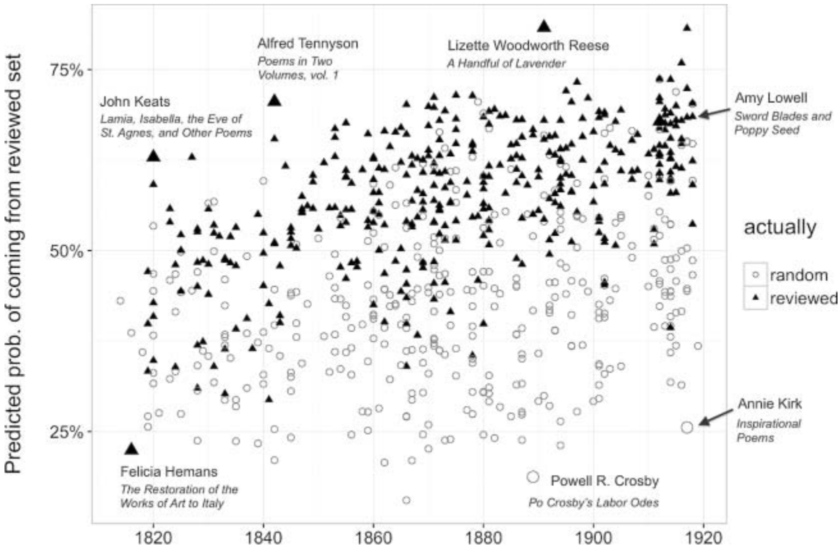

Figure 2. The same model, with outliers and interesting cases labeled

of the first-person plural in general.[9] 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) (fig. 2). A patient observer could tease out many other details.[10] For instance, feminine pronouns are also more common in reviewed works; if there are many of them (as in a poem with central female characters), they become strong evidence of literary prominence.

It may seem scandalous that statistical models can predict poetic reception without paying attention to versification or rhyme. The rise of *vers libre* normally plays an enormous role in our narrative of this period. Of course, free verse has that central role partly because critics disagreed about it, so it's not clear that it would be useful as a predictive

[9] June Howard (1999: 73) 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.

[10] For an interpretation of poetic prestige that categorizes words linguistically, see Kao and Jurafsky 2015. We deliberately trade linguistic categories for close reading here, mindful that our readers will be mostly literary historians.

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 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 critical disagreement, it foregrounds a relatively stable dimension of aesthetic reception—a dimension 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.* De Windt (1845: 10) uses rhyme, and "the *mountains* / Frowned in *majesty sublime*" may ostensibly be "romantic," but the model places Keats (1820: 110) much closer to Whitman (1872: 29). It is 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, we find it hard 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 advantage of distant reading is that it can be more patient with historicism, revealing even slow changes as historical phenomena. The criteria of judgment revealed by a model of reception cannot be interpreted as "an objective means of ranking" poems (Dalvean 2015), because patterns of reception do change over time. But they change quite slowly, at a pace that might be difficult to distinguish from permanence using our ordinary critical toolkit.

How Quickly Does Reception Change?

This article is in some ways a continuation of 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 part of their results. In "Slaughterhouse" Moretti (2000: 221–23) sets out to discover changes in the logic of plot across a single decade—and ends by speculating that those changes were probably diffused across a longer timeline. Here we set out to model standards of poetic distinction in several twenty-year periods and find that it makes more sense to model a century as a single unit. In short, it is beginning to look as if received narratives of literary history generally lead scholars to overestimate the pace of change.

We do not claim to have produced an all-purpose metric for literary change here: that might be a challenging task. Measuring change in reception is a more tractable problem, because it involves a boundary between groups of texts. If our models predict that boundary reliably, we know that they have captured something important about reception; if one model can predict the boundary reasonably well across a century, we know that some important aspects of reception changed slowly. Poetry itself may have changed in other ways: D. H. Lawrence writes things about the sex lives of whales that would have made Alfred Tennyson blush. But that change will do little to alter a model of reception unless there was some period between 1820 and 1919 when erotic descriptions of whales started to make the difference between literary success and obscurity. In practice, the textual differences associated with success seem to have changed slowly.

To be confident on this point, we have compared multiple models. The model represented above tries to find a pattern that can explain a whole century at once. It is significant that it succeeds, but that is 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 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 rest of the century. That actually works: models trained only on a quarter century of the evidence are still right (on average) about 76.8 percent of the volumes in the whole data set (fig. 3). This gives us a rough-and-ready answer to the question posed in this section: how quickly does reception change? At the boundary we model here (getting reviewed in selective periodicals), standards changed quite slowly. None of these models can explain reception perfectly, because reception is shaped by all kinds of social factors, and accidents, that are not legible in the text. But a significant chunk of poetic reception can be explained by the text itself (the text supports predictions that are right almost 80 percent of the time), and that aspect of poetic reception remained mostly stable across a century.

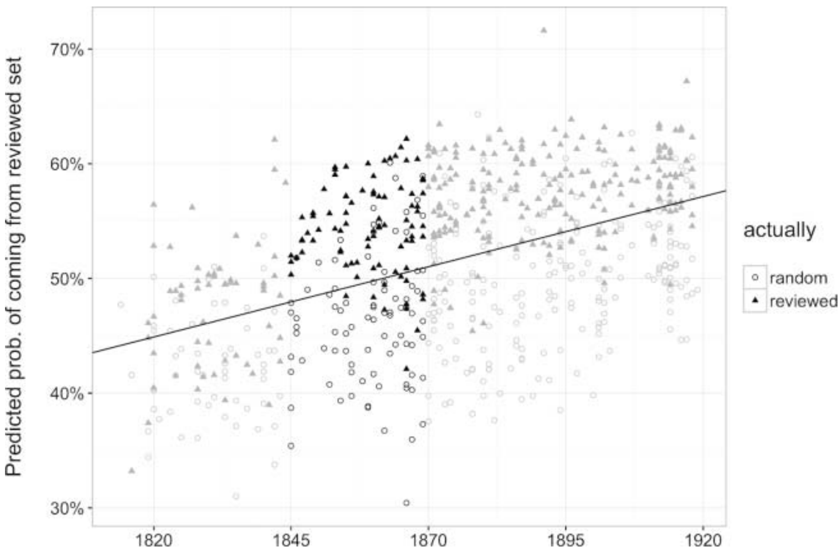

Figure 3. A model trained only on volumes from 1845 to 1869 makes predictions about the whole century

Synchronic Distinction and Diachronic Change

Nevertheless there were many changes, some of them even visible in the figures above. In all these models, the median probability that a volume will be reviewed appears to increase across time. That is not literally true. Reviewed and random volumes are evenly distributed across the timeline, so the probability of review remains constant. But the words common in reviewed volumes (across the whole timeline) also tend to become more common in *all* volumes as we approach the end of the timeline, so the cloud of data points always tilts upward. We have compensated for this by allowing the dividing line between categories to tilt upward as well. Although the standard of review worthiness remains the same, it is in effect applied more exactingly as time passes: poetic diction has to be ever darker, ever more concrete, to actually cross the bar.

This pattern is durable: 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 compels this chronological pattern to appear. We do not see a strongly marked, consistent tilt if we predict other social boundaries, like genre or authorial gender. But our work in progress on reviewing standards in fiction does reveal a similar diachronic pattern. This evidence has led us to hypothesize a general relationship between literary distinction and historical change. Diachronic change across a period seems to recapitulate the period's synchronic axis of distinction.

A conjecture this broad needs a few provisos. We don't yet know with certainty that this will happen outside the period 1820–1919; we are just hypothesizing that it will. And this is not 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 is more than one way to model distinction. Here we have focused on the outer boundaries of literary attention, but other scholars might emphasize distinctions closer to the center of the spotlight (say, prizes)[11]—and those might produce a different model.[12] So it is not as though the whole sweep of literary history has to move in any single direction forever. We just suggest that, whenever scholars do define a linguistic proxy for social distinction in a given period, they will find that change *relative to that axis* moves in an upward direction during the period itself. This pattern isn't shocking: it is easy to imagine reasons why it might happen. But that is not to say that we expected it or already understand it.

It is actually odd that a model trained on 1845–69 sees works from the 1870s as more likely to be reviewed than the works it was trained on. We didn't expect to see this, and we don't want to claim that we understand 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. In that case, 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 formed

[11] For the ambiguous status of nineteenth-century prizes, see English 2005: 28–49.

[12] We suspect, however, that different models of distinction will be mostly congruent. Although the model presented here is based on a binary contrast (reviewed or not), hypercanonical writers like 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.

their criteria of literary distinction partly by contrasting "the latest thing" to "the embarrassing past." In fact, causal arrows could point in both of these directions.

There are ways for us to untangle this causal knot. It is interesting, for instance, that predicted probabilities of review correlate with authors' dates of birth more strongly than they correlate with publication dates.[13] But as social scientists understand all too well, causal processes are hard to trace in detail. 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 have presented already suggests that some things scholars tend 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.

Gender and Nationality

The methodology we are using is close to social science, and we need to be alert for the interactions between variables that preoccupy social scientists. For instance, if women were less likely to be reviewed, our model might confound literary prestige with masculinity. Its predictions about reception would seem accurate only because it was leaning on the depressingly reliable assumption that works by women don't get reviewed.

To diagnose problems like this, we have recorded biographical information about authors and checked for interactions in a wide range of ways. In the case of gender, we have not seen an interaction that would distort the results described above. Women are underrepresented in this data set, contributing only about a quarter of the works overall.[14] 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 we run the modeling process on a

[13] The significance of this pattern is illuminated by Moretti 2000: 222 and Schmidt 2011.

[14] Women may have been even more underrepresented among poetry reviewers (Shattock 2007: 381).

data set restricted to women, it works just as well. 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 make poets more successful. (On the other hand, feminine pronouns may become slightly less significant as a positive force; they seem, perversely, to help men more than women.)[15] In short, gender certainly changes the boundaries of poetic distinction in ways worth studying, but we see no evidence that it undermines the broad conclusions drawn here.

The question of nationality is more vexed, because we draw 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 Francisco or Cincinnati than from Leeds. 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 percent of the time, but writers from the United Kingdom and Ireland only 74.7 percent 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 have observed.

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" (Gray 2001: 347). 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.

[15] For hints toward an explanation of this paradox, see Christ 1987.

Our original plan was to compare four twenty-year periods. We had planned to begin in 1840, because we didn't expect the style associated with elite taste to be clearly distinct in the period 1840–59 (perhaps a model would be only 60 percent accurate). But as we proceeded toward the twentieth century, our model would presumably get more and more accurate as the styles aimed at different reading audiences became clearly differentiated. We hoped for 80 percent accuracy by the twentieth century—which is what we found there. But what we didn't find was a very significant blurring of boundaries in earlier periods. Over most of the century we are now studying, there was only a slight change. Our data divide into two equal-size parts in 1876. A model limited to volumes published before that year is 77.3 percent accurate; one limited to volumes from 1876 forward is 80.5 percent accurate. This slight change may not even be a real signal, since the rising proportion of American authors in the collection plays a confounding role. It is certainly a less dramatic change than we expected when we began. Instead of "an increasing tendency throughout the nineteenth century for poetry to become a discourse of distinctly high culture" (Riede 1994: 445), we see an elite poetic culture that is already strongly differentiated from other poetic production by the mid-nineteenth century.

Because this result perplexed us, we pushed our start date back to 1820, and it turns out that if you reach back far enough, the differentiation of poetic styles does start to blur. Figure 1 shows that reviewed and random volumes are more evenly mixed before 1840; accuracy in that section of the timeline is only 66.7 percent. Accuracy remains low if you try modeling these early volumes by themselves. In other words, the early part of the timeline is not just organized by a stylistic boundary different from the one established later; there really appears to be less consensus about stylistic prestige in the first twenty years. There are reasons to phrase this conclusion cautiously (we had fewer sources for reviews in this period, and a lot of the reviewed volumes that are hard to classify come from one journal, the *Quarterly Review*), but it does seem that stylistic differentiation was weaker before 1840 than it later became.

Even for later decades received wisdom about the emergence of a "great divide" may be misleading only when it fails to specify which side of the divide was transformed more deeply. As book historians have recognized, the new literary institutions that arose 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., the "Invasion of the Tinsel Rhymesters" described in Newcomb 2004). These forms of differentiation were not 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 already stable by the second half of the nineteenth century). When we address fiction, we plan to use collections of best sellers and pulp fiction to explore differentiations within the broad part of the literary field that is here reduced to a single random sample.

Conclusion

A lot of descriptive work remains to be done in literary history, because we still know relatively little about patterns above the scale of a few hundred books. Literary historians have often generalized about the pace of change, for instance—contrasting epochs of relative stability to the "revolutions" that separate them (Fallis 1976; Greenblatt et al. 2006: 1834). But those claims are based on limited evidence. Scholars have supported theses about revolution by pointing to manifestos that crystallize important critical debates and to the differences between a few celebrated works. This makes sense if those texts constituted a vanguard, establishing patterns that rapidly swept the rest of the literary field. But did they? It has been difficult to say, because our existing critical methods struggle to describe changes in the literary field at large. It is particularly hard to describe changes in standards of reception, since claims of that kind entail comparisons between the relative fortunes of many works.

In this article we have shown that predictive models can trace the boundary between works consecrated by publicity and those that remained obscure. In Anglo-American poetry, the implicit standards governing reception had coalesced by 1840 and changed only slowly through 1919. This is not necessarily to say that poetry itself changed slowly. An enormous amount of ink has been spilled, persuasively, about the changes that distinguish Romantics from Pre-Raphaelites, and Pre-Raphaelites from modernists. The evidence presented here doesn't make that story unimportant, but it does suggest that structural changes in the literary field constitute a different narrative. Literary historians cannot keep assuming that the standards governing literary reception at large are a slightly delayed echo of generational conflict at the elite end of the field.

We haven't yet explained why the gradient of poetic prestige took the particular form we find in this period (promoting poems with concrete diction, relatively dark subject matter, and female characters; ignoring poems that are abstract, celebratory, or written in the first-person plural). That question will require more discussion, especially since our preliminary work on the reception of fiction reveals similar patterns. The explanation we need may have to cover not just poetry but nineteenth-century literary prestige more generally.

That question is too big for this article, but precisely because it covers so much ground, we should expect many parts of the answer to be found in existing scholarship. Reviewers' coolness to uplifting rhetoric, for instance, might be explained in part as a reaction against sentimentality (Howard 1999). Where gender is concerned, Carol Christ (1987: 385) notes, "the literary becomes increasingly feminized in the eighteenth and nineteenth centuries." The growing prestige of concrete description might overlap with what we describe elsewhere as the nineteenth-century emergence of "a specialized literary language" that defines itself against mere social capital, and especially against learned and abstract diction (Underwood and Sellers 2012). But these speculations still leave fascinating details unexplained—for instance, the odd preference nineteenth-century reviewers display for things that are blind, hollow, dull, or blank.[16] Literary scholars haven't described the history of reception very fully yet; it shouldn't be surprising that we cannot yet fully explain it.

The methods presented in this article may be unfamiliar in literary studies, but they are not fundamentally at odds with more familiar forms of interpretation. Scholars can model a social boundary statistically, for instance, but also interpret that boundary by looking closely at the literary pleasures that flourished on either side. We have only had space to do that briefly here; readers who want to see more case studies can consult a longer working paper online (Underwood and Sellers 2015).

[16] Some clues about the poetic value of blankness and indifference might be found in Pierre Bourdieu's (1995: 60–68, 77–81) discussion of Baudelaire.

Distant reading conflicts with literary historians' existing goals only in the practical sense that it takes a lot of time. This article discusses a medium-size corpus of 720 volumes, but to get those texts, we had to map a much larger digital library (including nearly a million volumes). Like other researchers in this field, we share our code and data, and we hope over time that practice will reduce bibliographic and technical obstacles to a project like this (Underwood 2015). But distant reading will continue to take time; figuring out what it really can or cannot do will require patience.

Ted Underwood is professor of English at the University of Illinois at Urbana-Champaign and author, most recently, of *Why Literary Periods Mattered: Historical Contrast and the Prestige of English Studies* (2013). His essay "Historical Difference as Immortality in the Mid-Nineteenth-Century Novel" appeared in the December 2002 issue of *MLQ*.

Jordan Sellers is instructor of English at the University of Illinois, Urbana-Champaign. His most recent essay, "Zero Ground: Mapping Maritime Commemoration in the Age of Nelson," appeared in the December 2015 edition of *European Romantic Review*.

References

Abbott, Andrew. 1995. "Things of Boundaries." *Social Research* 62, no. 4: 857–82.

Algee-Hewitt, Mark, Sarah Allison, Marissa Gemma, Ryan Heuser, Franco Moretti, and Hannah Walser. 2016. *Canon/Archive: Large-Scale Dynamics in the Literary Field*. January. Stanford Literary Lab Pamphlet Series. litlab.stanford.edu/LiteraryLabPamphlet11.

Algee-Hewitt, Mark, and Mark McGurl. 2015. *Between Canon and Corpus: Six Perspectives on Twentieth-Century Novels*. January. Stanford Literary Lab Pamphlet Series. litlab.stanford.edu/LiteraryLabPamphlet8.

Bourdieu, Pierre. 1995. *The Rules of Art: Genesis and Structure of the Literary Field*, translated by Susan Emanuel. Stanford, CA: Stanford University Press.

Breiman, Leo. 2001. "Statistical Modeling: The Two Cultures." *Statistical Science* 16, no. 3: 199–231.

Christ, Carol. 1987. "The Feminine Subject in Victorian Poetry." *ELH* 54, no. 2: 385–401.

Dalvean, Michael. 2015. "Ranking Contemporary American Poems." *Digital Scholarship in the Humanities* 30, no. 1: 6–19.

De Windt, Caroline. 1845. *Melzinga: A Souvenir*. New York.

DeWitt, Anne. 2015. "Advances in the Visualization of Data: The Network of Genre in the Victorian Periodical Press." *Victorian Periodicals Review* 48, no. 2: 161–82.

English, James F. 2005. *The Economy of Prestige: Prizes, Awards, and the Circulation of Cultural Value.* Cambridge, MA: Harvard University Press.

Fallis, Richard. 1976. "Yeats and the Reinterpretation of Victorian Poetry." *Victorian Poetry* 14, no. 2: 89–100.

Goldstone, Andrew. 2016. "Of Literary Standards and Logistic Regression: A Reproduction." January 4. andrewgoldstone.com/blog/2016/01/04/standards.

Gray, Janet. 2001. "Popular Poetry." In *Encyclopedia of American Poetry: The Nineteenth Century,* edited by Eric Haralson, 347–49. New York: Routledge.

Greenblatt, Stephen, et al., eds. 2006. *Norton Anthology of English Literature.* 8th ed. Vol 2. New York: Norton.

Heuser, Ryan, and Long Le-Khac. 2012. *A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method.* May. Stanford Literary Lab Pamphlet Series. litlab.stanford.edu/LiteraryLabPamphlet4.

Howard, June. 1999. "What Is Sentimentality?" *American Literary History* 11, no. 1: 63–81.

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

Kao, Justine T., and Dan Jurafsky. 2015. "A Computational Analysis of Poetic Style: Imagism and Its Influence on Modern Professional and Amateur Poetry." *Linguistic Issues in Language Technology* 12, no. 3: 1–31.

Keats, John. 1820. *Lamia, Isabella, the Eve of St. Agnes, and Other Poems.* London.

Long, Hoyt, and Richard Jean So. 2016. "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning." *Critical Inquiry* 42, no. 2: 235–67.

Loomis, George. 1891. *Memorial or Decoration Day: A Poem Founded on Facts.* N.p. catalog.hathitrust.org/Record/009558911.

Mason, Nicholas. 2013. *Literary Advertising and the Shaping of British Romanticism.* Baltimore, MD: Johns Hopkins University Press.

Moretti, Franco. 2000. "The Slaughterhouse of Literature." *MLQ* 61, no. 1: 207–27.

———. 2005. *Graphs, Maps, Trees.* London: Verso.

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

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30.

Riede, David G. 1994. "The Victorian Era." In *Columbia History of British Poetry,* edited by Carl Woodring and James S. Shapiro, 425–51. New York: Columbia University Press.

Rossetti, Christina. 1865. “Echo.” In *Goblin Market and Other Poems*. 2nd ed., 81–82. London.

Schmidt, Benjamin. 2011. “Age Cohort and Vocabulary Use.” Sapping Attention, April 11. sappingattention.blogspot.com/2011/04/age-cohort-and-vocabulary-use.html.

Sculley, D., and Bradley M. Pasanek. 2008. “Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities.” *Literary and Linguistic Computing* 23, no. 4: 409–24.

Shattock, Joanne. 2007. “Reviewing.” In *A Companion to Victorian Poetry*, edited by Richard Cronin, Alison Chapman, and Antony H. Harrison, 378–91. Oxford: Wiley-Blackwell.

Shmuéli, Galit. 2010. “To Explain or to Predict?” *Statistical Science* 25, no. 3: 289–310.

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

Underwood, Ted. 2014. “Page-Level Genre Metadata for English-Language Volumes in HathiTrust, 1700–1922.” Figshare. December 29. DOI:10.6084/m9.figshare.1279201.

———. 2015. “Paceofchange.” Code and data used to support the present article, as well as Underwood and Sellers 2015. GitHub repository. DOI:10.5281/zenodo.44226.

Underwood, Ted, and Jordan Sellers. 2012. “The Emergence of Literary Diction.” *Journal of Digital Humanities* 1, no. 2. journalofdigitalhumanities.org/1-2/the-emergence-of-literary-diction-by-ted-underwood-and-jordan-sellers.

———. 2015. “How Quickly Do Literary Standards Change?” Figshare. May 19. DOI:10.6084/m9.figshare.1418394.

Whitman, Walt. 1872. *Leaves of Grass*. Washington, DC.

Wickham, Hadley. 2009. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.