Machine Learning and Human Perspective

TED UNDERWOOD

CONTROVERSIES ABOUT THE MEASUREMENT OF LITERATURE ARE much older than the phrase *digital humanities*. Many of the questions now pursued with computers can be traced back to researchers who worked "with pen and graph paper" in the middle of the twentieth century (Buurma and Heffernan). Objections to quantitative methods can be traced back even further. For more than a century, critics have observed that numbers stabilize neutral facts about the past at the cost of flattening out the perspectival variety that gives history meaning. In 1910, the perspectival character of human experience was Wilhelm Dilthey's reason for separating the human sciences from the natural sciences (154–64). In 1973, Stanley Fish objected to the quantitative study of literary style for a similar reason. Computers may tell us that Jonathan Swift used a lot of connective words. But the literary meaning of Swift's style, Fish argued, could vary from one interpretive community to another.

These objections never stopped literary scholars from using numbers in certain special contexts, such as book history. But exceptions of that kind only seemed to prove the rule that intellectual life was broadly divided between "interpretive" disciplines and "quantitative" ones (Smith). Contemporary debates about digital humanities may appear to retrace the same divide. Franco Moretti presents distant reading as a project that replaces "interpretation" with "the explanation of general structures" (91). Critics have responded that quantitative analysis will always miss the point of literature, which is to foreground the "infinite variation" of "human perception" (Drucker, "Why" 634). Both sides seem to be restating the twentieth-century consensus that numbers are useful for measuring objective facts but not for interpreting perspectival differences. Scholars who already know that their

TED UNDERWOOD is professor of information sciences and English at the University of Illinois, Urbana. He is the author of three books, most recently *Distant Horizons: Digital Evidence and Literary Change* (U of Chicago P, 2019).

© 2020 TED UNDERWOOD
PMLA 135.1 (2020), published by the Modern Language Association of America

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

own work is interpretive may understandably feel that they can skip the whole debate.

For most of the twentieth century, that was a safe policy. But the boundary between quantitative and interpretive methods was always permeable, and recent intellectual advances have made it easy to traverse. Since learning algorithms rely on examples rather than fixed definitions, they can be used to model the tacit assumptions shared by particular communities of production or reception. This approach gives quantitative research a new flexibility, which is allowing scholars to survey culture from specific vantage points in the past and even to measure the parallax between vantage points. Literary historians have only recently begun to exploit these possibilities. But it is already clear that our inherited assumptions about the difference between measurement and interpretation need to be revised. In order to understand the shape of emerging theoretical debates, we will need to explore the perspectival uses of machine learning.

From Measurement to Modeling

Many things have changed since Fish wrote his critique of quantitative literary analysis in 1973. Computers are faster now; digital libraries are bigger. Since these changes of scale are easy to understand, digital humanists often use them to explain growing reliance on numbers in the humanities. Critics are not wrong to feel that this explanation is incomplete. To be sure, numbers tend to be more useful at large scales of analysis. But scale has done nothing in itself to resolve the hermeneutic problem that troubled earlier forms of computational research. The concepts that interest literary historians are still hard to measure, in part because their meanings change from one context to another.

To understand why quantitative research has nevertheless made real progress in recent decades, we need to dig beyond hype (and alarm) about big data and notice subtle shifts of strategy. One of those shifts has been described in recent *PMLA* articles that bracket debates about computers and the canon in order to refocus theoretical conversation on the concept of a "model" (So; Piper). Instead of trying to measure a stable concept, humanists who build statistical models typically study a boundary between two social contexts. They try to understand this social boundary by measuring differences associated with it.[1] The differences measured may not be important in themselves; they could involve things as trivial as punctuation marks. The goal of the inquiry is not to measure anything with inherent significance but rather to define a model—a *relation* between measurements—whose significance will come from social context.

Consider the problem of representing gender. Whether we understand gender as a performance (Butler) or as a real "position one occupies" (Alcoff 148), it is clear that gender is relational. It is less a fact about the subject than about the subject's relation to a social audience. So gender categories change their meaning as we move from one context to another. Masculinity may not have meant the same thing in 1950 that it did in 1850, and it may not mean the same thing for women that it does for men. Categories with this relational character are best represented indirectly. It would be fruitless to define or measure masculinity, since the category means very little in itself. But the transformations of masculinity—as we move from one period or perspective to another—are a topic that a model could illuminate.

A recent article I wrote with David Bamman and Sabrina Lee explores the transformations of gender by comparing the language used in characterization over a span of two centuries (Underwood et al., "Transformation"). A program called BookNLP clusters the names of the same character—so "Maggie Tulliver" and "Maggie" can be treated as the same person in *The Mill on the Floss*—

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

and then identifies words grammatically associated with each person. This approach gives us only a fraction of the insight a human reader might extract. In the following passage, the only words the program links to Maggie are the ones I have italicized:

> The resolute din, the unresting motion of the great stones, giving her a dim delicious *awe* as at the presence of an uncontrollable force—the meal forever pouring, pouring—the fine white powder softening all surfaces, and making the very spider-nets look like a faery lace-work—the sweet pure scent of the meal—all helped to make Maggie *feel* that the mill was a little world apart from her outside every-day *life*. The spiders were especially a subject of speculation with her. She *wondered* if they had any relatives. . . . (Eliot 1: 45–46)

Arguably, George Eliot is using the flowing, tumbling energy of the mill, hidden from the outside world, to characterize Maggie metonymically. BookNLP misses that metonymy and knows only that Maggie is someone who has "awe" and a "life" and who "feels" and "wonders." But that isn't a terrible summary of the broad strokes this passage uses to characterize her. Even if we are capturing only broad strokes, interesting patterns become legible when we swing BookNLP across 87,800 books and millions of characters in English-language fiction (Underwood et al., "Replication Data").

For instance, we can ask how the signs of gender used in characterization vary with authors' identities. One could eventually pose that question about many different authorial roles—including collective authors and pseudonyms, as well as cis and trans identities. But we might start by considering perspectives on gender expressed by writers who publicly identified as either "men" or "women." (Public identification is not expressed only on title pages; for example, Eliot eventually stepped forward as a woman, and she is recorded as one here.) Both axes in figure 1 measure the tendency for a word to be associated grammatically with feminine or masculine characters. Each axis is, in short, a model of a particular perspective on gender. The only difference between the axes is that the vertical axis measures gender differentiation in books by women and the horizontal axis measures it in books by men. We can think of the image as a model of the relation between two perspectives.

The northeast and southwest corners of figure 1 contain words whose gendered connotations are a subject of broad agreement. No matter who wrote the book, feminine characters tend to have mothers and hair and say "oh." Masculine characters tend to have beards and pockets and say "sir." Things get more interesting when we look along the other diagonal.[2] In the northwest corner, we find words that men tend to apply to men and that women tend to apply to women. It turns out that when Maggie "wondered" about spiders, she was doing something authors associate with characters of their own gender. The same thing holds true for remembering, thinking, hearing, and seeing. These verbs are clearly signs of subjectivity. It is less obvious why authors claim certain body parts (the feet, throat, head, and stomach) for their own gender identity.

In the southeast corner, we find a group of words that men use in describing women and that women use in describing men. Heteronormative patterns are visible here; most authors talk more about marrying, kissing, and loving when a character doesn't share their gender identity. Passive roles are also prominent. I have used the prefix "was-" to indicate cases where a character is the object of a verb rather than its subject. So the sentence "In the darkness of that night she *saw* Stephen's face turned towards her in passionate, reproachful misery" counts as an instance of "saw" for Maggie but as an instance of "was-seen" for Stephen Guest (Eliot 3: 231; emphasis added). Other words in this region describe how a character was seen, even when the character

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

is the subject of the sentence: they describe a character's "expression" or "tone," or how the character "seemed" to another observer. It is clear, in short, that women writers tend to describe men externally—and vice versa.

This pattern is not shocking. But that is not to say we knew it all in advance. And some details remain unclear. I still don't know why writers claim the stomach and feet for their own gender. More importantly, it would have been plausible to expect subjectivity and passivity to be gendered primarily in books by men—while books by women might distribute seeing and being seen equally among their masculine and feminine characters. In many other cases, the distorting effects of gender are concentrated among men. For instance, men create lopsided ensembles of characters, rarely more than a third of them women, while women tend to balance their dramatis personae evenly (Underwood et al., "Transformation"). But the perspectival distortions traced in figure 1 are largely symmetrical, with blind spots on both sides. Now that we have traced this symmetry, it may appear, in hindsight, inevitable. But it isn't something we actually knew, or something we fully understand even now.

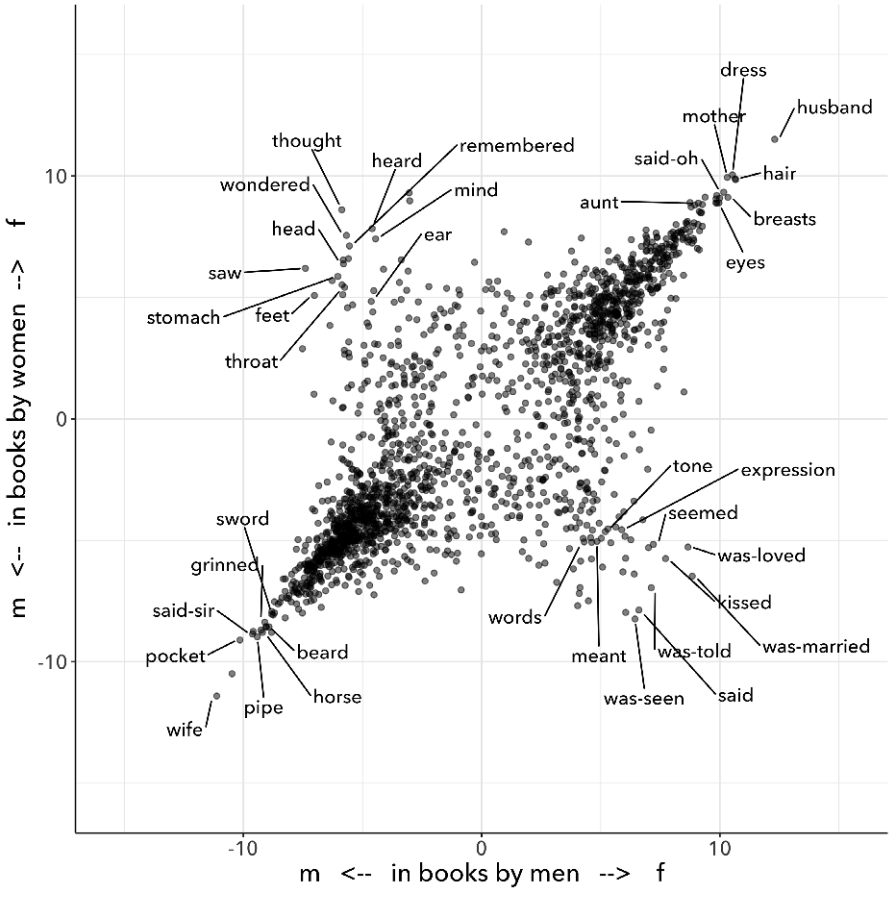


FIG. 1

The gendering of words used in characterization. The data set is composed of 87,800 works of English-language fiction published between 1780 and 2007. On each axis, positive numbers indicate that a word is overrepresented in descriptions of women and negative numbers indicate that it is overrepresented in descriptions of men. The scale is the signed log of the log-likelihood ratio (Dunning). Areas with many overlapping dots look darker.

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

I have offered figure 1 as an example of the indirection of contemporary quantitative methods. Instead of trying to define gender, this diagram models a contrast between two perspectives on the topic. I recognize that gender theorists will be frustrated by the binary structure of the diagram. To be sure, this binary has folded back on itself, in order to acknowledge that social systems look different from different positions in the system. But the diagram does still reduce the complex reality of gender identification to two public roles: men and women. I needed a simple picture, frankly, in order to explain how a quantitative model can be said to represent a perspective. But nothing about this method compels us to stop at two perspectives. We can also multiply gender identities, pose intersectional questions, or inquire about historical transformations of gender, as my coauthors and I recently attempted (Underwood et al., "Transformation").

Miriam Posner has rightly characterized the fluidity of digital ontologies as "the radical, unrealized potential of digital humanities." Making that potential into reality will require many strategies; the rest of this essay explores a particularly flexible one. Instead of turning two roles into four, or ten, we could jettison the whole premise that researchers must decide in advance on a fixed set of categories. Because figure 1 is plotting numbers loosely analogous to proportions (the representation of A relative to B), it compelled us to sort characters into mutually exclusive groups. And when groups are understood as exclusive, a researcher does need to draw up a list of categories in advance. But there are more flexible ways to build and compare models. The predictive models produced by machine learning don't require a world of exclusive categories with consistent definitions. All they require is an observer who can point at examples of something they have in mind. There is no limit to the number of observers.

Multiplying Perspectives

To explain how such a loose-jointed approach could work, the rest of this essay explores the history of genre—a topic that benefits enormously from flexibility. Critics once imagined genres as a limited set of natural literary kinds, each organized by a unifying rationale, like the concept of "the novum" that Darko Suvin thought unified works of science fiction (79). But late-twentieth-century historicism undermined that confidence (Warhol). Since the 1980s, scholars have tended to envision genres as "empirical, not logical" categories—"groupings [that] arise at particular historical moments," that "need not have a single trait in common," and that "are subject to repeated redefinitions or abandonment" as social conditions change (Cohen 210).

This theory of genre is appealing because it seems to promise a more flexible literary history, rooted in the imperfect continuities of human life rather than in imaginary universals. But the theory doesn't provide a way to measure degrees of similarity between groupings created at different moments. So literary historians still often fall back on lumping and splitting strategies of an all-or-nothing kind. For instance, contemporary fans may see works by the nineteenth-century writers Mary Shelley and Jules Verne as examples of science fiction. But the phrase *science fiction* first appeared in the 1920s. Nineteenth-century readers didn't necessarily assume that Shelley's dark stories belonged with Verne's *voyages extraordinaires* ("extraordinary journeys") or with dream visions of a utopian future. Toward the end of the century, a concept of so-called scientific romance did begin to take shape around Verne and H. G. Wells. But some scholars distinguish scientific romance from science fiction and argue that the latter crystallized only in the second quarter of the twentieth century (Stableford; Westfahl). Others argue that neither concept is stable. According to one recent history, "there is no

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

such *thing* as SF—but instead multiple and constantly shifting ways of producing, marketing, distributing, consuming and understanding texts as SF" (Bould and Vint 1).

Frustration with the semantic character of these debates probably unites everyone who has participated in them. At this point, scholars know that genres don't really have crisp boundaries. We know that our task is not to define terms (or utterly reject them) but to trace the gradual mutations of social practice. Unfortunately, in tracing those changes it is hard to rule out the possibility that the continuities we perceive have been created by the very genealogical assumptions we set out to test. Stepping outside our own assumptions is always difficult, and especially difficult for historians because time only flows in one direction. To correct retrospective bias, a cautious researcher might like to send a box of twenty-first-century books back to readers in 1895 along with a letter asking them to pick out anything that looks like a scientific romance. Alas, a box of that kind is hard to send.

But perhaps we could use the documents dead people have left behind to reconstruct their practices of selection? Computer science may help. The point of machine learning is exactly to model practices of categorization that lack a definition and can be inferred only from examples. We may not know how to define spam, although we recognize it in our inbox. So a spam filter begins with a training set composed of messages that readers did or did not mark as spam. An algorithm learns to model spam using whichever textual details do in practice distinguish those groups of messages. It is a flexible strategy that excels at reproducing human behavior, but also a risky strategy when neutrality is the goal. A bank shouldn't use machine learning to winnow loan applications unless it plans to accept all the assumptions made by the people who approved or rejected loans in the training set. Even if overt signs of race and gender are removed from the data, an algorithm trying to reproduce human choices may well find proxies for race and gender buried in addresses and occupations.

But capturing the unfair, ill-defined assumptions implicit in a particular set of human choices is exactly what historians need to do. Literary historians know that the practices they want to reconstruct were not neutral or objective. For example, we know that the gender of the author may play a role when readers are categorizing a book as science fiction or fantasy. Our goal in modeling genre is not to rewrite history as if it had been fair but to represent real practices of selection so that we can trace degrees of similarity among the perspectives of different places and times. Strange as it may seem, machine learning is precisely suited to this purpose.

How does it work? The algorithms used in this article are supervised, which means that they learn from texts that have been labeled by human readers. Each perspective on genre is represented as a fuzzy boundary separating texts that were, or were not, assigned a particular genre label. If texts were points in three-dimensional space, this boundary would be a plane tilted at an angle that separated most of the works labeled fantasy from most of the ones with other labels. I say "most" because all statistical models are imperfect. We will be interested in the errors they make. But before inquiring about error, we need to ask how texts can be represented as points in space at all. What variables would count as height or width? Researchers could try to define variables cleverly suited to a particular genre. For instance, to identify fantasy, we might ask, How much does this plot depend on magic? But that would entail risky assumptions, since we don't really know that magic is essential to fantasy or how to distinguish magic from "sufficiently advanced technology" (Clarke 21n1).

Remember, however, that a model of loan applications didn't need explicit references to gender or race in order to absorb human bias. Similarly, although we may not know

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

which attributes define fantasy for a given reader, we can expect many of those attributes to leave traces somewhere in the text. So a model might simply count words, treating the relative frequency of each word as a dimension like height or width. Since the vocabulary of fiction contains thousands of words, this will produce a space with thousands of dimensions, but a space of that kind can still be divided by a hyperplane.[3]

A skeptic might protest that genres are defined not just by diction but by plot, setting, and theme—which is true. But this article is not aiming to define genres. Rather, it uses models to represent genres as practices of reception so that we can compare the practices of different eras. Models represent reception in an active and concrete sense, by re-creating the selection practices of a particular reader or group of readers. A model trained on a subset of texts labeled by a reader should be able to recognize not only the texts it was trained on, but other texts that the same reader assigned to the same genre. (In fact, to avoid circularity, models are always tested on these more challenging held-out examples, not the examples in their training set.) And while diction might not provide a very satisfying abstract definition of a genre, it is more than sufficient to support this kind of concrete re-creation. For example, statistical models based simply on words and punctuation can distinguish mysteries from works in other genres with 93% accuracy. If we peer into the internal workings of these models, we will find that they rely heavily on question marks and words expressing uncertainty, like *whoever*. While these details are technically matters of diction, they are also clearly shaped by the larger interrogative structure of a mystery plot. It turns out that many formal patterns leave verbal traces of this kind. So researchers have found that they don't need to represent plot and character directly in order to predict human judgments shaped by plot and character (Allison et al.; Underwood, *Distant Horizons* 34–67).

To illustrate the fluidity that quantitative models can add to our histories of genre, the rest of this article will compare a range of perspectives on science fiction and fantasy.[4] These genres are usefully puzzling. Their histories have been written in many different ways and are sometimes even collapsed into a single story about speculative fiction. A book-length study of this history might draw evidence from dozens of bibliographies or thousands of book reviews, and might explore differences between national traditions. This article sets out more modestly to illustrate the potential of a new approach and uses a more restricted kind of evidence. I will focus on works originally written in English (plus a few influential works in translation) and will compare only a handful of perspectives.

The most important source is the library itself. Genre classifications drawn from HathiTrust and OCLC allowed me to find thousands of volumes that librarians had labeled "fantasy fiction" and "science fiction."[5] But librarians' genre labels were almost all assigned in the last forty years. When attached to books published before 1980, they might be projecting a recent perspective on literary practices that were understood differently at the time. So I have also sought out earlier sources—a dissertation written in 1934, a circulating library catalog from 1911, and several early critical studies and bibliographies. Finally, I created a random background set composed of six hundred volumes of fiction from HathiTrust, excluding those marked as science fiction or fantasy. Combining all these sources gives us 1,581 volumes, dated in most cases by first publication. A collection of 1,581 volumes doesn't by any means exhaustively cover the history of genre. But the notion that quantitative inquiry aims at exhaustiveness has been oversold both by its recent advocates and by its critics. In truth, distant readers are always working with samples and are often more interested in modeling the differences between samples than in making claims about the whole library.

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

So what can perspectival models teach us about science fiction or fantasy? The words that end up predicting genres raise fascinating questions. Fantasy, for instance, can often be recognized by the words *tale*, *sunlight*, and *seven*. But I will mention individual words only briefly here, because this essay is not trying to provide a single, stable definition of science fiction or fantasy. Instead it compares multiple models to explore perspectival questions about the history of genre. Researchers have found that the genres easiest for computers to model are also the genres human readers tend to agree about (Calvo Tello). So comparing the strength of different models may cast new light on critical arguments about the relative stability of genres.

Let us start with models defined by the perspective of the librarians who assigned genre labels in the last forty years. It turns out that the volumes librarians call "science fiction"—sampled evenly from 1870 to 2010 with a few examples from earlier in the nineteenth century—can be recognized by a single model with 89.9% accuracy. This is surprising, since critics don't necessarily agree that works written before 1920 are really science fiction. Moreover, we tend to imagine that the distinguishing feature of this genre is technology, and technology changes quickly. Few of the inventions mentioned by Verne remain science-fictional today. But a linguistic model of this genre focuses less on submarines or rocket ships than on a general rhetoric of sublimity and ambiguity marked by large numbers, deliberately vague nouns like *thing* and *creature*, and verbs like *blink* and *groped*. Since this rhetoric unites authors ranging from Shelley to Kim Stanley Robinson, arguments that "there is no such *thing* as science fiction" (Bould and Vint 1) would appear to have exaggerated the genre's mutability.

With its anchor in the past, fantasy might seem more stable than science fiction: swords don't date as quickly as ray guns. But when we group all the volumes contemporary librarians call "fantasy," they can be identified only 84.5% of the time. The difference from the accuracy of the model for science fiction (89.9%) may not sound huge, but the gap between the genres gets bigger as we go back in time (fig. 2). This widening gap suggests that the boundaries of science fiction solidified earlier than the boundaries of the genre we now call fantasy. In fact, "fantasy" may not even be the right label to apply to works published before 1900; accuracy in that period falls low enough (77%) that one may wonder whether librarians who assigned the label "fantasy" were using an anachronistic concept.

The gap between science fiction and fantasy isn't the only interesting pattern in figure 2. The upward trend from 1875 to 1985 implies that both genres became easier to separate from other fiction in the library—although, in the case of science fiction, the change is only subtle. The increasing accuracy of these models is open to several explanations. We could be seeing the consolidation of generic conventions associated with plot or theme: a more crisply delimited genre might become easier to identify. But since the genre labels used in figure 2 were assigned recently, it is also possible (at least in the case of fantasy) that the labels are simply a better fit for late-twentieth-century literature than they are for literature from earlier periods.

The surprising part of figure 2 isn't that genres become easier to recognize but that the change is so subtle in the case of science fiction—which is often said not to have existed as a coherent genre before a crystallizing moment in the first half of the twentieth century. The creation of the magazine *Amazing Stories* in 1926 underwrites one popular origin story, because the magazine's emergence coincided with the emergence of the term *science fiction* itself (Westfahl 12). Gary K. Wolfe delays the moment of consolidation even longer, contending that "the science fiction novel persistently failed to cohere as a genre" until Pocket Books gave it form in the early 1940s

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

(21). But the fluctuations in figure 2 don't look like the emergence of a new genre. Instead, science fiction seems to be rather coherent already in the era of Verne and Wells—nearly as coherent as it is today, and more so than fantasy is today. The boundaries of the genre do become slightly clearer for a while in the second half of the twentieth century, but this is hardly a picture of a genre that "failed to cohere" until a first seed crystal took shape in 1926.

In fact, the blurring of boundaries in the last thirty-five years is at least as striking as any consolidation of science fiction before that point. Having raised doubts about Wolfe's origin story for science fiction, I should acknowledge that this recent trend does closely fit the thesis of his book *Evaporating Genres*. As Wolfe notes, the hybridity of science fiction and fantasy—both with each other and with the literary mainstream—has recently given rise to genre concepts like "Slipstream," "Bizarro," and "the New Weird" (164). He argues that these movements are symptoms of a more general diffusion. "Fantasy is evaporating . . . growing more diffuse, leaching out into the air around it, imparting a strange smell to the literary atmosphere" (viii). The downward turn for both genres at the end of the timeline in figure 2 supports his story.

I have been talking about *science fiction* and *fantasy* as if the terms applied equally to works in every period, although the evidence already hints that this may not be true for *fantasy*. One advantage of a perspectival approach is that we don't have to take continuity on faith. Observers at different points on the

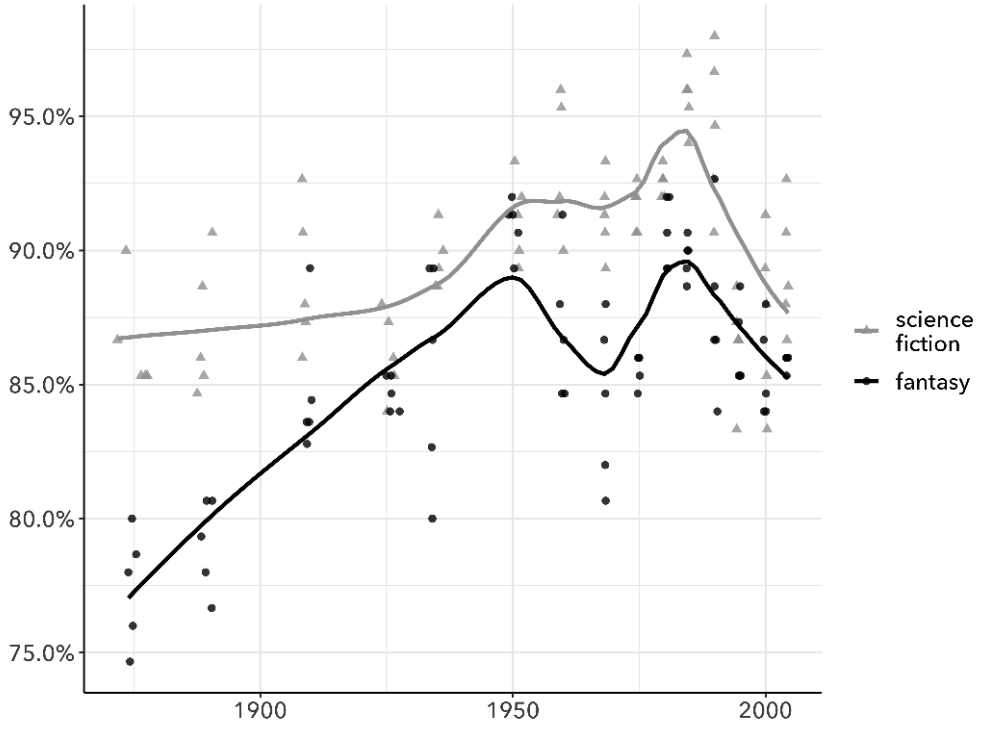


FIG. 2

The accuracy of models that identify volumes labeled "fantasy" or "science fiction" by librarians. Each point represents a single model trained on a random sample of texts from a certain period. Each model covers one hundred fifty texts; points are plotted at the mean publication date for those texts. Trend lines are drawn through the points by (an arbitrary amount of) locally weighted smoothing and should be taken with a grain of salt.

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

timeline could really be describing different things. To find out, we can model different perspectives and compare them.

In the case of fantasy, it may be difficult even to know which earlier perspectives to choose, because fantasy has connections to, among other things, children's literature and Victorian medievalism. Here I have space to trace only one possible genealogy, running back to fiction that early-twentieth-century readers chararacterized as "supernatural" or "occult." Dorothy Scarborough's critical study *The Supernatural in Modern English Fiction* (1917) mentions several authors often cited today as prototypes of fantasy, such as Lord Dunsany and William Morris. But Scarborough also dwells on many books we might not consider genre fiction, like Eliot's *The Lifted Veil*. Although this collection of works may look heterogenous to our eyes, we can test its underlying similarity to modern categories by running a version of the thought experiment that asks readers in the past and present to sort the same box of books. If we train a model on Scarborough's supernatural fiction and ask it to identify nineteenth-century works labeled "fantasy" by librarians in the last forty years, the model trained on Scarborough's collection is only 5% less accurate than a model trained on recent labels. Clearly there is some continuity between Scarborough's concept of the supernatural and our concept of fantasy.

The continuity doesn't prove, however, that our concept of fantasy provides a good description of nineteenth-century fiction. On the contrary, there is strong evidence that Edwardian categories are better at organizing the period. Figure 2 shows that our concept of fantasy doesn't correspond to clear boundaries between nineteenth-century works; a model trained on recent labels achieves only 77% accuracy. A model based on Scarborough's critical study does slightly better (81%), and best of all is a model based on categories implied by Edwardian marketing. For instance, we get 87% accuracy modeling the works of fiction gathered under the heading "Occultism" in Mudie's circulating library catalog for 1911 (*Catalogue* 886; see 886–88).

This triumph for categories framed by immediate contemporaries is roughly what recent genre theory might lead us to expect. The turn toward a historical conception of genre has led many theorists to conclude that the genre categories organizing a period are, by definition, whatever contemporary observers said they were (Rieder 191–92). The premise that genres are subject to frequent redefinition has also tended to shrink the range of observers who count as immediate contemporaries. Moretti, for instance, has conjectured that genres are really generational phenomena lasting for only "25–30 years" (21). Genres that seem to last longer might just be names that have been loosely applied to a sequence of distinct generation-sized phenomena. We already have some reason to reject Moretti's conjecture, since we have seen that it is relatively easy for a single textual model to recognize the generic kinship of works spread across two centuries. But comparing different perspectives on supernatural, occult, and fantasy fiction still appears to confirm the emphasis on contemporary observers in recent genre theory. The most accurate models seem to be produced by marketing categories from the time the books were written.

The story becomes very different, however, when we turn to science fiction. No version of this genre looms at all large in Edwardian marketing. The catalog of Mudie's circulating library places five "pseudo-scientific" novels in a subcategory of a section on the "mysterious and marvellous," but this is a tiny detail in the catalog, comparable to the four books grouped as "snake mysteries" within the much larger section on "occultism" (*Catalogue* 886). We have nevertheless seen that books retrospectively labeled "science fiction" by librarians do in fact compose a recognizable division of nineteenth-century literary

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

practice. Although this category is a retrospective projection, it can be modeled just as accurately as the categories actually listed in contemporary catalogs and indexes. Moreover, we have reason to believe that the patterns organizing this category remained relatively stable from the nineteenth century to the present, in spite of several name changes. I have already mentioned that a single model can identify with a high degree of accuracy books drawn from any point on the two-century timeline. For a more severe test of continuity, we can break the timeline in half and compare the halves. I trained one model on books mentioned in *Scientific Fiction in English, 1817–1914*, a dissertation written by James O. Bailey in 1934, before "science fiction" had become a widely accepted term. The other model was trained on books from 1915 to 1975 that were labeled "science fiction" by librarians after 1980. When I asked each model to sort the other model's list of books, I got an average accuracy of 78%. This would not be impressive for a model trained and tested on the same period, like the models used to create figure 2, but since these models were trained on works selected by different observers using slightly different terms to characterize different centuries, I would call it a significant degree of continuity. Similar tests on fantasy never achieve the same degree of stability, even if we allow contemporary librarians to select works in both halves of the timeline.

In short, a literary practice closely comparable to twentieth- and twenty-first-century science fiction did exist in the nineteenth century, although contemporary observers paid it scant attention and gave it a range of different names when they noticed it at all. Admittedly, this evidence of historical continuity reverses the conclusion we might expect a perspectival method to produce. I started by assuming that observers from different eras were describing different objects, only to discover that—in the case of the *voyage extraordinaire*, scientific romance, etc.—these different objects were bound together more tenaciously than many scholars have believed. The experiment did begin with perspectival premises. But a well-designed experiment can challenge its own premises.

I have juxtaposed the divergent stories of science fiction on the one hand and supernatural, occult, and fantasy fiction on the other in order to suggest that genre theory needs a more flexible framework than our present habits of argument can give it. Formalism taught us to believe that genres are durable implicit categories. Historicism is teaching us to believe that they are explicitly defined by contemporary observers. Neither theory is always reliable, because the word *genre* can cover a wide range of phenomena—patterns that last a decade or several centuries, overt marketing strategies or literary practices that acquire a name only in retrospect. Critics are not unaware of this complexity. We have struggled to acknowledge it in several ways—for instance, by distinguishing genres from the looser patterns we call "modes." But even this distinction is probably too coarse. Many of the transactions that puzzle literary historians take place in a no-man's-land between genre and mode. The Gothic arguably began as a genre and diffused outward—at some point—to become a mode (Richter 161–62). Moreover, as figure 2 shows, even practices understood as genres can have dramatically different degrees of stability.

To describe the full variety of patterns in literary history, we need a descriptive language that can acknowledge differences of degree. In acknowledging those gradations, numbers don't confine critical description; they rather liberate it from a fixed taxonomy. I have shown that fantasy and science fiction can be modeled separately. But they are also sometimes combined to create a larger tradition called speculative fiction. Quantitative methods don't need to stall out here in a semantic debate between lumpers and splitters. Instead we can accept the reality of any practice readers actually

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

recognize and simply measure the distance between different practices. It turns out that models of science fiction lose only 9 to 11% accuracy when asked to recognize fantasy (and vice versa). So these genres are closer to each other than either is to detective fiction (where the models would lose 30% accuracy), but not quite as close as, say, fantasy is to Scarborough's supernatural fiction. In the last thirty years fantasy and science fiction have grown even closer. A model of one now loses only 6% accuracy when asked to identify the other.[6]

Measuring Parallax

Our picture of genre has been sketched in broad strokes so far, but we can also use numbers to fill in color and detail. Since the models discussed above make predictions about individual works, it becomes easy to ask which works typify a particular perspective—or a particular contrast between perspectives. For instance, if we want to understand the changes that separate prewar from wartime and postwar science fiction, we can train a model on examples of SF published from 1910 to 1939. Then we can ask our prewar model to identify works of science fiction published in the next thirty years (1940–69) and compare its predictions to the predictions of a model trained directly on the later genre. This tells us which examples of later science fiction are most surprising from a prewar perspective—hardest for a model trained on prewar works to recognize as science fiction. Since all these volumes were tagged as SF by postwar librarians, we are no longer contrasting the selection practices of different readers. Having found a great deal of agreement among readers of science fiction, we are now contrasting versions of science fiction defined by the practices of writers in different periods. Figure 3 visualizes the contrast. Think of each arrow as a measurement of parallax, revealing how a particular book seems to shift position as we move from a prewar to a postwar vantage point.

Long upward-pointing arrows suggest that a work was very different from prewar SF. Many of these titles are legible as emblems of generic change. Judith Merril is celebrated as one of "one of the most visible—and voluble—apostles of the New Wave in 1960s sf;" it makes sense that the anthologies she edited would dramatize new trends (Latham 251). Robert Heinlein's *Stranger in a Strange Land* is also notorious for posing social questions that would have been alien to *Amazing Stories*. However, Jack Vance's picaresque novella *The Dying Earth* is slightly easier to recognize as science fiction from a vantage point in the past.

The Dying Earth may have been old-fashioned; it is set in a stylized distant future and lacks the relative social realism of much postwar SF. But before speculating about the explanation for the downward motion of the tiny arrow in figure 3, we should ask how much evidence it actually conveys. The starting and ending points of each arrow represent the average of thirty different modeling runs. Each run uses a slightly different sample of books to model prewar and postwar perspectives on science fiction, and each run assigns different probabilities to individual volumes. The shaded circles for *Player Piano* represent a typical range of variation for one book: roughly 68% of probability estimates fall somewhere within the shaded circles. Even a casual visual comparison suggests that the average difference between perspectives on *The Dying Earth* will be small relative to random variation. A statistical test confirms that the difference between prewar and postwar perspectives on Vance's novella could easily occur by chance. The other four measurements of parallax labeled in figure 3 do all represent statistically significant changes. But measurements of parallax are rough approximations and cannot by themselves tell us *how* a particular book diverged from earlier examples of a genre. To interpret this evidence, we need to supplement it with more familiar forms of critical inquiry.

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

In fact, figure 3 is just a slightly systematized version of a thought experiment that literary historians often attempt. If we want to understand why Jane Austen's *Emma* was important, we situate ourselves imaginatively in a world without Henry James and Eliot—a world where the nearest point of comparison is Maria Edgeworth. Perspectival models won't replace that kind of imaginative immersion; they aren't nearly as sensitive as a human reader. They do, however, have the advantage of genuine ignorance. It can be difficult for a human being with a PhD to forget that the nineteenth century happened. But a model based on thirty years of evidence knows nothing beyond those thirty years. It doesn't have to fake amnesia, and this makes it a valuable informant—as if it were a visitor from 1815 or 1935 whose reaction to later works we could observe.

If we want to more fully understand a model's reaction, we can watch it read a text. For instance, we don't have to speculate about the aspects of Ursula Le Guin's *The Left Hand of Darkness* that might have stretched prewar definitions of science fiction. We can apply prewar and postwar models to individual passages from the book and look for places where they disagree. The passage where the models diverge most sharply happens to be discussing the book's central premise: that people on the planet Gethen can play different reproductive roles at different points in their lives. Prewar models don't see this passage as typical of science fiction. In what follows I have italicized the words

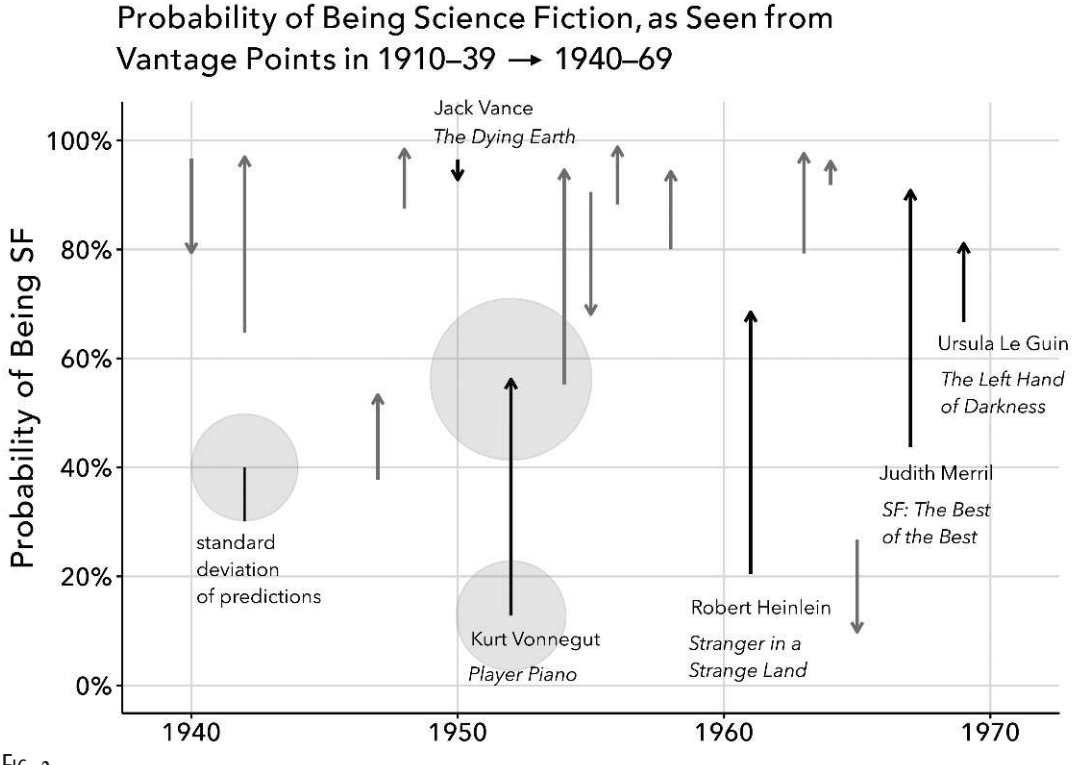


FIG. 3

Two different sets of models predict the probability that a volume is science fiction. The start of each arrow is the probability perceived by models trained on examples from 1910 to 1939; the end is the probability perceived by models trained on examples from 1940 to 1969. Labeled books are in black; those not labeled are left gray. The vertical radii of the shaded circles for *Player Piano* represent the standard deviation of predictions for that book; predictions for other books vary in a similar way.

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

that contribute most strongly to the difference of opinion:

> Consider: Anyone can turn his hand to anything. This sounds very *simple* but its *psychological effects* are incalculable. The fact that everyone between seventeen and thirty-five or so is liable to be (as Nim put it) "tied down to childbearing" implies that no one is quite so thoroughly "tied down" here as women elsewhere are likely to be psychologically or *physically. Burden* and *privilege* are *shared* out pretty equally; everybody has the same risk to run or *choice* to make. Therefore nobody here is quite so *free* as a *free male anywhere* else. (93–94)

The word *male* does thematize sexuality. But that is just one of the ways this passage signals a change in the science fiction genre. More broadly, its language is shaped by psychological and social reasoning (*burden*, *privilege*, *free*, *choice*, *shared*, *psychological effects*). Heinlein's politics were rather different from Le Guin's, but a close (model-assisted) reading of *Stranger in a Strange Land* suggests that it baffles prewar models of science fiction in a similar way: by telling a story that hinges on psychological and social conflict rather than on the solution to a physics problem.

In short, quantitative evidence can be relevant at any scale—from a paragraph to a book to a century-spanning trend. But it may have different degrees of importance at different scales. At the paragraph level, quantitative signals are often swamped by noise. For example, the italicized words above are a small sample from a longer book, loosely illustrative rather than statistically significant. More crucially, our familiar reading strategies are just too strong at this scale to need a lot of help. Readers already know that Le Guin turned from the physical sciences to the social sciences. The paragraph I have quoted from *The Left Hand of Darkness* comes from a chapter that presents itself as the "field notes" of an anthropologist (89). The point of reading this chapter over a model's shoulder is not really to confirm our critical intuition but to confirm the usefulness of the model.

As we back up to a larger scale of description, more surprising patterns become visible. For instance, it hasn't always been obvious that Heinlein was diverging from generic tradition in the same broad way as Le Guin. Zooming out even farther, we start to glimpse patterns that were invisible at the ordinary scale of reading. Measuring the distance between models trained on different periods, we discover that science fiction has often changed slowly (more slowly than fantasy). The contrast described above, across a dividing line at 1940, takes place in the period of most rapid change. Science fiction changed more across this divide than it did across the 1920s, when the term itself was invented. This is not a conclusion that will be immediately intuitive to every historian, so even distant readers need to build bridges between scales, showing how abstract measurements of historical distance are related to the innovations visible on particular pages.

The history of genre traced above diverges from received opinion in at least two ways. I have downplayed the transformation wrought by *Amazing Stories* and emphasized the long-term stability of scientific romance and science fiction compared with the things now called fantasy. But I have not been arguing that numbers give these conclusions any special authority. Statistical models are just one more form of evidence, to be weighed along with all the others. This article recommends numbers to literary scholars not as a uniquely reliable form of evidence but as a flexible descriptive language especially suited to historicist questions of perspective and of degree.

In theory, to be sure, we already know that genres are contingent, blurry constructions. But our descriptive vocabulary still tempts us to draw crisp boundaries. We say that the Gothic is a genre, or merely a mode, or a genre that at some point became a mode. We say that fantasy is distinct from science

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

fiction or isn't, or became fully distinct only after *The Lord of the Rings*. The strategy I have called "perspectival modeling" can support a more flexible approach to description, which begins with the practices of historically situated readers and traces degrees of affinity among them. Literary constellations will take shape as perspectives from different periods recognize each other and are pulled together. The map that emerges from this process may still have some familiar landmarks—for instance, a blurry region labeled "scientific romance" next to an even blurrier region with overlapping labels like "fantasy" and "supernatural fiction." But this map is a continuum: instead of arguing about semantic boundaries, historians can use it to measure degrees of contestation and relative speeds of change.

Computational Hermeneutics

Critics of quantitative approaches to literature often argue that numbers imply an objectivity incompatible with what Johanna Drucker calls the "relativistic and comparative methods of the humanities" ("Humanistic Theory"). Drucker has advanced this case most systematically, arguing that the quantitative methods we share with other disciplines make evidence seem "self-evident, value neutral, and observer-independent" ("Humanities Approaches"). But Drucker is not alone; this critique of numbers is so common that it can be reduced to a dismissive gesture. For example, Alexander R. Galloway implies in passing that "bean counting" is for "young [professors] who don't understand hermeneutics."

This article has argued that quantitative methods are well-suited to comparative, relativistic, hermeneutic questions. Numbers don't inherently promise objectivity; they are just signs invented by human beings to reason about differences of degree. Since numbers quickly became useful in natural science, we have learned to associate them with physical measurements that don't depend greatly on an observer's position. But that is hardly the limit of their usefulness. Today, even natural scientists are using statistics to acknowledge researchers' subjective assumptions, or priors. Humanists can take the lead in using statistical models to represent the conflicting perspectives of historically situated observers.

Critics of computation are right to warn researchers against treating computers as objective oracles. Unsupervised algorithms in particular are often granted more authority than they deserve. The unsupervised algorithms used for clustering and topic modeling don't require examples labeled by human beings. Just pour in texts and patterns come out. Occasionally, researchers imagine that the absence of direct human supervision gives these results a special authority. Alan Liu has called this fantasy "*tabula rasa* interpretation" (414). In reality, even unsupervised algorithms are designed by human beings who make assumptions about the patterns they expect to find.

In the hands of writers who understand those assumptions, unsupervised algorithms have valid uses. But this essay has explored a different approach, which grounds interpretation more explicitly in human history. Supervised machine learning closely resembles humanists' traditional approach to the past, using documents produced in another place and time to reconstruct a vanished perspective on the world. The chief difference between supervised models and more familiar methods is that a supervised model can make predictions about new evidence—predictions that allow the model to behave like a living observer and make revealing mistakes. By studying those mistakes, we can map the parallax between perspectives and measure differences of degree that would be hard to represent without numbers.

To address the common assumption that numbers inherently posit objectivity, I have spent much of my energy in this article showing that machine learning can excel at

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

answering slippery perspectival questions. But I have also used more familiar methods, from close reading to reflection on Edwardian marketing. Since different approaches are suited to different questions, there is, of course, no right way to interpret literature. Sometimes we need to closely examine one paragraph from *The Left Hand of Darkness*; sometimes we need to compare a series of models stretching across two centuries. Instead of trying to resolve methodological conflict by crafting a compromise, this article has explored a wide range of interpretive options that it presents as compatible and connected.[7] "The more options the better" has been my implicit premise.

If quantitative methods don't conflict with humanists' existing theories of interpretation, why has controversy on the topic been so fierce? I believe the real problem is that new methods remain inaccessible. Since students of literature are not trained in statistics, they don't see methods that rely on numbers as opportunities meant for them. Instead, new methods look like an opportunity for others—a prospect that rarely brings unmixed joy to the human heart. That would remain true even if researchers could use a learning algorithm to refract the remembered taste of a madeleine into a synaesthetic rainbow that appeared different to every observer. A method becomes humanistic not when it has the right philosophical character but when it is in fact used by humanists.

Digital approaches to the humanities have been accepted most readily when they promise to package new methods as user-friendly tools. But for complex questions, a user-friendly interface can rarely be more than a gateway drug. To test their own conclusions, literary scholars who use numbers will need knowledge of statistics and a bit of programming experience. If we integrate those subjects in the literary curriculum, our discipline may soon find itself exploring change and continuity in new perspectival ways. If we don't, no argument will be eloquent enough to make new methods popular. A discipline not trained to use statistical models will necessarily see them as foreign competition. In that case, questions like the ones explored in this article will probably be asked and answered by social scientists instead.

Notes

I was able to write this article because a sabbatical from the University of Illinois and a Meyer H. Abrams Fellowship at the National Humanities Center freed me from other duties.

1. For the importance of "boundaries" in this approach to culture, see Abbott.

2. I used a logarithmic scale to make this diagonal of disagreement more visible; if I had used an ordinary scale, the diagonal would have been dwarfed by agreement about gender.

3. These models were created using regularized logistic regression from scikit-learn (Pedregosa et al.). Features included the most common words and punctuation marks, along with a few stylistic features like sentence length; data was drawn from Capitanu et al. The number of features in each model was selected by grid search, as was the regularization constant. Visualizations were produced using ggplot2 (Wickham). See Underwood, "Code" for more details.

4. The idea of comparing fantasy and science fiction came from Alan Liu.

5. At a later stage of inquiry it might be worthwhile to separate national traditions, but works by Jules Verne and Karel Čapek are perfectly legible as science fiction, even when translated. I also regularize British and American spelling, in order to pose questions that span the Atlantic. Children's literature has been excluded from this analysis, because its strong association with fantasy raises questions that would require a longer study.

6. For simplicity's sake, I measure the distance between genres (and later on, the pace of change within a genre) by measuring the accuracy lost when a model trained on one group of works tries to recognize the boundary defining a different group. There are more precise ways to compare models. I also simplify by assuming that the distance from model A to model B is effectively the same as the distance from B to A. These simplifications don't significantly distort the conclusions reported here, but they do play down the non-Euclidean geometry of culture. For an experiment that measures the divergence between models more precisely and without Euclidean assumptions, see Underwood, "Historical Significance."

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

7. Literary scholars use numbers in a wider range of ways than a single article could explore. I haven't said much about models that represent geography (Evans and Wilkens) or cross linguistic boundaries (Long and So). Nor have I fully explored the methodological debates that divide scholars who use numbers (Bode; Goldstone; Klein; Tenen). I have, admittedly, implied a general response to those debates by suggesting that different approaches are usually compatible. Corpora can be compared; scales of analysis can be connected.

Works Cited

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

Alcoff, Linda Martín. *Visible Identities*. Oxford UP, 2006.

Allison, Sarah, et al. "Quantitative Formalism: An Experiment." *Stanford Literary Lab*, 15 Jan. 2011, litlab.stanford.edu/LiteraryLabPamphlet1.pdf.

Bailey, James O. *Scientific Fiction in English, 1817–1914: A Study in Trends and Forms*. 1934. U of North Carolina, PhD dissertation.

Bode, Katherine. "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History." *Modern Language Quarterly*, vol. 78, no. 1, 2016, pp. 77–106.

Bould, Mark, and Sherryl Vint. *The Routledge Concise History of Science Fiction*. Routledge, 2011.

Butler, Judith. *Gender Trouble: Feminism and the Subversion of Identity*. Routledge, 1990.

Buurma, Rachel Sagner, and Laura Heffernan. "Search and Replace: Josephine Miles and the Origins of Distant Reading." *Modernism/Modernity*, 11 Apr. 2018, modernismmodernity.org/forums/posts/search-and-replace.

Calvo Tello, José. "Genre Classification in Spanish Novels: A Hard Task for Humans and Machines?" European Association for Digital Humanities, Galway, Dec. 2018. *EADH 2018*, eadh2018.exordo.com/files/papers/46/final_draft/20181205_genre_classification_human_vs_machines.pdf.

Capitanu, Boris, et al. The HathiTrust Research Center Extracted Feature Dataset 1.0. *HathiTrust Research Center*, 2016, doi:10.13012/J8X63JT3.

Catalogue of the Principal English Books in Circulation at Mudie's Select Library. Mudie's Select Library, 1911.

Clarke, Arthur C. *Profiles of the Future: An Inquiry into the Limits of the Possible*. Rev. ed., Harper and Row, 1973.

Cohen, Ralph. "History and Genre." *New Literary History*, vol. 17, no. 2, 1986, pp. 203–18.

Dilthey, Wilhelm. *The Formation of the Historical World in the Human Sciences*. Princeton UP, 2002.

Drucker, Johanna. "Humanistic Theory and Digital Scholarship." *Debates in the Digital Humanities*, edited by Matthew K. Gold, U of Minnesota P, 2012, dhdebates.gc.cuny.edu/debates/text/34.

———. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly*, vol. 5, no. 1, 2011, www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

———. "Why Distant Reading Isn't." *PMLA*, vol. 132, no. 3, May 2017, pp. 628–35.

Dunning, Ted. "Accurate Methods for the Statistics of Surprise and Coincidence." *Computational Linguistics*, vol. 19, no. 1, 1993, pp. 61–74.

Eliot, George. *The Mill on the Floss*. 3 vols., William Blackwood, 1860.

Evans, Elizabeth F., and Matthew Wilkens. "Nation, Ethnicity, and the Geography of British Fiction." *Journal of Cultural Analytics*, 13 July 2018, culturalanalytics.org/2018/07/nation-ethnicity-and-the-geography-of-british-fiction-1880–1940/.

Fish, Stanley. "What Is Stylistics and Why Are They Saying Such Terrible Things about It?" *Approaches to Poetics*, edited by Seymour Chatman, Columbia UP, 1973, pp. 109–52.

Galloway, Alexander R. "Everything Is Computational." *Los Angeles Review of Books*, 27 June 2013, lareviewofbooks.org/article/franco-morettis-distant-reading-a-symposium/.

Goldstone, Andrew. "The *Doxa* of Reading." *PMLA*, vol. 132, no. 3, May 2017, pp. 636–42.

Klein, Lauren F. "Distant Reading after Moretti." *Arcade*, 2019, arcade.stanford.edu/blogs/distant-reading-after-moretti.

Latham, Rob. "Sextrapolation in New Wave Science Fiction." *Science Fiction Studies*, vol. 33, no. 2, July 2006, pp. 251–74.

Le Guin, Ursula. *The Left Hand of Darkness*. Ace, 1986.

Liu, Alan. "The Meaning of the Digital Humanities." *PMLA*, vol. 128, no. 2, Mar. 2013, pp. 409–23.

Long, Hoyt, and Richard J. So. "Turbulent Flow: A Computational Model of World Literature." *Modern Language Quarterly*, vol. 77, no. 3, Sept. 2016, pp. 345–67.

Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for Literary History*. Verso, 2005.

Pedregosa, Fabian, et al. "Scikit-learn: Machine Learning in Python." *JMLR*, vol. 12, 2011, pp. 2825–30.

Piper, Andrew. "Think Small: On Literary Modeling." *PMLA*, vol. 132, no. 3, May 2017, pp. 651–58.

Posner, Miriam. "What's Next: The Radical, Unrealized Potential of Digital Humanities." *Miriam Posner*, 27 July 2015, miriamposner.com/blog/whats-next-the-radical-unrealized-potential-of-digital-humanities/.

Richter, David. *The Progress of Romance: Literary Historiography and the Gothic Novel*. Ohio State UP, 1996.

Rieder, John. "On Defining SF, or Not: Genre Theory, SF, and History." *Science Fiction Studies*, vol. 37, no. 2, July 2010, pp. 191–209.

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press

Scarborough, Dorothy. *The Supernatural in Modern English Fiction*. G. P. Putnam, 1917.

Smith, John K. "Quantitative versus Interpretive: The Problem of Conducting Social Inquiry." *New Directions for Program Evaluation*, no. 19, 1983, pp. 27–51.

So, Richard J. "All Models Are Wrong." *PMLA*, vol. 132, no. 3, May 2017, pp. 668–73.

Stableford, Brian. *Scientific Romance in Britain, 1890–1950*. Fourth Estate, 1985.

Suvin, Darko. *Metamorphoses of Science Fiction: On the Poetics and History of a Literary Genre*. Edited by Gerry Canavan, Peter Lang, 2016.

Tenen, Dennis Yi. "Toward a Computational Archaeology of Fictional Space." *New Literary History*, vol. 49, no. 1, 2018, pp. 119–47.

Underwood, Ted. "Code and Data to Support 'A Measured Perspective.'" *Zenodo*, 1 Apr. 2018, doi:10.5281/zenodo.1210966.

———. *Distant Horizons: Digital Evidence and Literary Change*. U of Chicago P, 2019.

———. "The Historical Significance of Textual Distances." *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Aug. 2018, pp. 60–69. *ACL Anthology*, aclweb.org/anthology/W18-4507.

Underwood, Ted, et al. "Replication Data for 'The Transformation of Gender.'" *Harvard Dataverse*, 2018, doi: 10.7910/DVN/ZM2MAN.

———. "The Transformation of Gender in English-Language Fiction." *Cultural Analytics*, 13 Feb. 2018, culturalanalytics.org/2018/02/the-transformation-of-gender-in-english-language-fiction/.

Warhol, Robyn. "Genre Regenerated." Introduction. *The Work of Genre: Selected Essays from the English Institute*, edited by Warhol, English Institute / American Council of Learned Societies, 2011, hdl.handle.net/2027/heb.90055.0001.001.

Westfahl, Gary. *Mechanics of Wonder: The Creation of the Idea of Science Fiction*. Liverpool UP, 1999.

Wickham, Hadley. *Ggplot2: Elegant Graphics for Data Analysis*. Springer, 2009.

Wolfe, Gary K. *Evaporating Genres: Essays on Fantastic Literature*. Wesleyan UP, 2011.

https://doi.org/10.1632/pmla.2020.135.1.92 Published online by Cambridge University Press