

Measuring Literature: Digital Humanities, Behavioral Economics, and the Problem of Data in Thomas Piketty's *Capital in the Twenty-first Century*



It was common in the enthusiastic reactions to the English translation of Thomas Piketty's *Capital in the Twenty-first Century* to mention his use of literature. David Harvey thought Piketty "spiced up" his data about the history of income equality "with neat literary allusions to Jane Austen and Balzac." Larry Summers notes that *Capital* is "littered with asides referencing Jane Austen and the works of Balzac." Paul Krugman thrilled that Piketty's book was "a work that melds grand historical sweep with painstaking data analysis," asking, "When was the last time you heard an economist invoke Jane Austen and Balzac?" Stephen Marche argues that Piketty's *Capital* could be "reasonably

mistaken for a work of literary criticism” and imagined future economic historians using the “socialist realist novels” of 2008’s Great Recession in the same way that Piketty uses Austen and Balzac.

As a literary scholar, I think we need to reevaluate such enthusiasm about Piketty’s use of literature as data. In particular, I believe literary critics need to assess the consequences of the feeling, shared among these reviewers, that literature can provide facile and transparent access to the economic realities of the nineteenth century. In this essay, I challenge this assumption by examining the complex ways in which Piketty deploys literature to undergird his assertions about the economic past of Europe. Then, I demonstrate how Piketty’s use of literature, while considered and reflective, originates in the attitudes of behavioral economics, despite his concerted attempts to avoid it. Finally, I turn to the analysis of literary scholars who examine Piketty’s use of literature using the techniques of “big data.” Their techniques show how the use of literature in Piketty’s *Capital* points to more fundamental differences about what qualifies as evidence in cultural study.



1. An engraving of Humphry Repton’s “Burley on the Hill” from *Observations on the Theory and Practice of Landscape Gardening, including Some Remarks on Grecian and Gothic Architecture* (London, 1803), p. 132. Repton was a successful landscape architect who used the country house at Burley as an example of how to preserve the natural gifts of scenery by compromising between “ancient and modern gardening, between art and nature.” Courtesy of Special Collections, D.H. Hill Library, North Carolina State University.

In *Capital*, Piketty uses literature in two ways that contrast quite strongly with those identified by his laudatory reviewers. First, he uses literary characters to personalize his arguments, making them more accessible to readers and portable for other criticism. He calls the insight that inheritance makes individuals richer than their income ever could “Vautrin’s Lesson.” He names the problem of inherited wealth dominating those who gain money from income “Rastignac’s Dilemma.” Both of these phrases draw on figures from the novels of

Balzac and both of these literary episodes help structure Piketty's book by serving as section titles. These titles reveal that some of Piketty's thinking about historical economics originated in literature. He said in a 2014 interview that he initially became interested in questions of wealth accumulation by wondering whether Rastignac's anxieties about becoming rich were widely shared in nineteenth-century France or were unique to Balzac who, Piketty notes, was "obsessed with his own debt."

In addition to personalizing larger economic forces, Piketty uses literature to illustrate the effects of historical economic realities and, more controversially, he sometimes uses literary references themselves as the evidence for that empirical reality. For example, what Piketty calls the "classical patrimonial society" of late eighteenth-century and early nineteenth-century Europe is also named "the world of Balzac and Austen." This world is typified by the sort of gentry country estate depicted in the works of the era's most famous landscaper, architect Humphry Repton, and by the manners of its inhabitants, represented in the novels of Austen.

The contrast to these lavish country estates that Repton helped create and that Austen described was the impoverished nineteenth-century world of urban France often found in Balzac's novels, whose moments are so frightening because, Piketty claims, they "[contain] such precise figures" itemizing its penury.



2. Frontispiece, *Pride and Prejudice*, by Jane Austen (London and New York, ca. 1890). Courtesy of the American Antiquarian Society, Worcester, Massachusetts. Austen's novels provide Piketty with an example of what he describes as "classical patrimonial society" and the social manners of its country gentry.

Piketty is well aware of the difficulty of economists' confidence that their specific models best represent economic realities. He offers that the "verisimilitude and evocative power" of novelists like Austen and Balzac is one

“no statistical or theoretical analysis can match.” “Film and literature, nineteenth-century novels especially,” he enthuses, are “full of detailed information about the relative wealth and living standards of different social groups.” Because the “physical reality of inequality” also possesses a “fundamentally subjective and psychological dimension,” Piketty believes there will always be a value in the way art and literature capture income inequality. The artistic reveals otherwise rarefied econometrics and presents critical rhetorical strategies for popularizing the abstruse findings of economics.

The question that Piketty never resolves—perhaps cannot be expected to resolve—is the consequence of seeing literature not just as “verisimilitude” but also as “detailed information,” as potentially another form of those distribution tables and income indexes that make up his ample and publicly available datasets. He insists, for example, that among Austen, Balzac, and their readers, “money had the same meaning.” The shared meaning of money resulted from the stability of monetary values and the consistent wealth accumulation that occurred as nations used taxes to pay off the public debts to its creditors. “Hence,” Piketty writes, “it is no surprise that wealth is ubiquitous [*omniprésent*] in Jane Austen’s novels: traditional landlords were joined by unprecedented numbers of governmental bondholders.” Piketty immediately continues: “(These were the same people, if literary sources count as reliable historical sources).” (In the French, he writes this assertion as: “...en grande partie les mêmes personnes, si l’on en croit les récits littéraires comme les sources historiques...”)

The language of this parenthetical phrase explicitly considers whether literary sources “count” (“can be believed” or “can be trusted”) as reliable historical sources of economic data. Despite the seeming uncertainty introduced by Piketty’s language, his analysis seems to indicate that he believes that literature does count—in every sense of that term—as reliable information. He argues, for example, that *Germinal* or *Oliver Twist* “did not spring from the imaginations of their authors, any more than did the laws limiting child labor.” Instead, he seems to propose that they arose from historical circumstances for which their authors are a kind of conduit. Balzac may have been obsessed with his debt, but it was the economic realities that propelled his writing and created the shared understandings among Austen’s readers.

From one vantage, this is not an especially controversial way to read literature; for decades, literary critics have evaluated how empirical reality is differentially represented in imaginative writing. But Piketty’s confidence in the ways literature can capture the empirical realities of the economic past shares much with the ideologies of behavioral economists like Dan Ariely and Tyler Cowen, who explain life events as disparate as house purchases, poetry readings, and food selection as decisions of taste and preference reducible to quantifiable forces of supply, demand, and price.

The tendency to explain social phenomena through economic models and quantification is not new, arguably dating back to Gary Becker, an innovator in

the field of the economics of human behavior. As Becker writes in his 1975 paper on money and marriage, economists can use “economic theory ... to explain behavior outside of the monetary market sector, and increasing numbers of noneconomists have been following their examples.” Becker describes marriage, for example, as a “scarce resource” in a market economy. Such scarcity has consequences for social organization, reproduction, and population growth, leading him to conclude that the marriage market demonstrates “compelling additional evidence on the unifying power of economic analysis.” Personal and socio-cultural choices have underlying economic dynamics, Becker concludes, whether individuals are aware of it or not.

Michel Foucault recognized the gravity of this shift toward economic analysis as a “unifying power.” Foucault referred explicitly to Becker in his 1970s lectures at the Collège de France as an origin of neoliberalism. He noted that the economic human being described in Becker’s research “appears precisely as someone manageable, someone who responds systematically to systematic modifications artificially introduced” so that he is “the correlate of a governmentality.” (Becker humorously responded to Foucault’s critique in 2012, claiming he “like[d] most of it and did not disagree with much.”)

Becker’s sense that economic analysis can be modified to evaluate any human behavior is a powerful methodology with consequences for how we think about art and literature. Consider this account from Dan Ariely, who, in an effort to prove that “we are all economists” who “hold the basic beliefs about human nature on which economics is built,” recalls an experiment he devised about the concept of price anchoring. Price anchoring is a notion that humans overly rely on an initial price when they determine the value of a good or service. Ariely describes how he begins his experiment by reading from Walt Whitman’s *Leaves of Grass* to a group of students. He then asks one group of students whether they would pay him \$10 to have him read poetry; he asks a separate group of students whether they would listen to him read poetry if he paid them \$10. Afterward, he solicited bids for his poetry reading services from all of the students. He found that those asked if they would pay him offered more money for poetry reading than those whom he offered to pay. The initial “anchor”—whether Ariely seemed ready to pay or be paid to read poetry—altered their monetary valuation of the same experience.

Of course, Ariely conflates this monetary valuation with the “pleasure” or “pain” of an aesthetic experience. He concludes that in his experiment he is like Tom Sawyer, for “[m]uch like Tom Sawyer, I was able to take an ambiguous experience [poetry reading] ... and arbitrarily make it into a pleasurable or painful experience” depending on the price—that is, depending on whether students thought they were getting a “good” price to hear Ariely read.

This example is meant to be partly humorous, as Ariely speaks with self-deprecation about his poor skill at reading poetry. But the intermixture of price anchoring with the performance of Whitman’s poetry and the economic interpretation of Twain are linked directly to the methodological confidence of

a figure like Becker. The primary assumption is that cultural experiences, like encountering Whitman's poetry, fundamentally depend on price; art, like iron ore, is a "scarce resource" whose quality can be quantified.

It is within this context that Piketty's arguments about the reliability of literature as an archive of historical economic data become so crucial. As literature becomes a dataset for econometrics, whether it is pursued by the methods of Piketty or Ariely, we are increasingly forced to ask ourselves what kind of data literature provides.

One answer might be offered by the new forms of literary criticism developing at the intersection of big data, digital humanities, and distant reading. Using these techniques, Ted Underwood, Hoyt Long, and Richard Jean So examined one of Piketty's assertions about literature and economics: the supposedly precipitous decline in novelistic references to money in the twentieth century. For Piketty, this amounts to the dissolution of money as possessing a shared meaning as it did in the "age of Austen and Balzac." Underwood, Long, and So disagree, concluding that while readers should "trust" Piketty on the significance of income inequality, they should "ignore what he says about literature." Piketty's account of literary history is "wrong," they claim; in fact, "it's exactly the reverse of Piketty's story about the disappearance of money" with references to "specific units" of currency in English-language literature nearly doubling (from 2 instances to 4 instances per 10,000 words of text) between 1800 and 1950, the period during which Piketty maintains it declines.

They reach these conclusions by identifying references to monetary values in 7,700 novels published between 1750 and 1950 found in HathiTrust Digital Library. However, as with most criticism, the dilemma of computational analysis is determining what qualifies as data to be put into the model. For digital humanists—and for literary critics especially—these decisions about data are provoked in part by the excursions of social scientists, especially economists, into the literary.

Is it significant, either culturally or formally, that the number of references to money nearly double between 1800 and 1950 in 7,700 novels? It's unclear whether the mathematics tell us that it is or isn't. As Underwood, Long, and So themselves suggest, such results might be explained by the changing audience of novels over these two centuries. Typically, these questions of significance have been resolved by matching measurable data—for example about changing patterns of literacy in the Anglo-American world—to assertions about how individual literary works are constructed and used by readers. For me, the literary text is the bedrock unit of analysis, and examining how it is constructed and influenced by historical and political forces is the basis of my professional analysis. I make an argument in concert with a corpus of primary and secondary texts to support the significance of my observations.

The quantitative analysis of the kind Underwood, Long, and So apply to

Piketty's assertions might compel a reexamination of this model by forcing literary scholars to use other measures of significance. It might be that these quantitative measures of literature reveal underlying patterns of significance that can only be viewed from extremely large gatherings of texts. Or perhaps, as with the analysis of the paragraph by Mark Algee-Hewitt, Ryan Heuser, and Franco Moretti, it is an attempt to sensitize us to overlooked (and undervalued) structures of literature. Much of the unease about the scholarship of digital humanities results from the way big data and digital humanities have inventively altered the form, especially the visual form, of literary criticism by populating it with graphs, tables, charts, and numerical figures about word frequency, word proximity, and topic modeling.

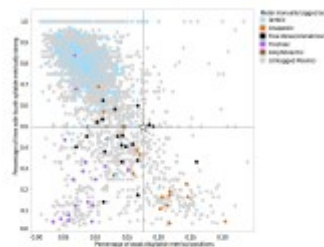


Figure 12. Mapping English Poetry
A plot of PROXIMITY (vertical axis) versus FREQUENCY (horizontal axis) for the subjects through the lexicon corpus. Only the subjects with a significant difference in the two axes are shown. The legend indicates the subjects: Drama, Novel, Non-fiction, Poem, and Unlabeled. The subjects are color-coded: Drama (blue), Novel (red), Non-fiction (green), Poem (orange), and Unlabeled (grey).

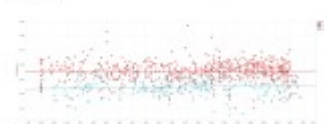


Figure 13. Cross-sectional visualization of word frequency
This figure shows the frequency with which combinations of any two consecutive words appear together in a sample of 1000 words. The x-axis represents the frequency of the first word, and the y-axis represents the frequency of the second word. The legend indicates the subjects: Drama, Novel, Non-fiction, Poem, and Unlabeled. The subjects are color-coded: Drama (blue), Novel (red), Non-fiction (green), Poem (orange), and Unlabeled (grey).

3. A page from “On Paragraphs: Scale, Themes, and Narrative Form” by Mark Algee-Hewitt, Ryan Heuser, and Franco Moretti (Stanford Literary Lab Pamphlet 10, October 2015) that demonstrates the enormous graphical diversity involved in literary criticism associated with digital humanities. Courtesy of the Stanford Literary Lab and the Stanford University Libraries.

For a literary criticism that has been dominated by a single form—continuous prose occasionally interrupted and accented by representational images—these changes in form are substantial and should not be overlooked. They require literary critics to read arguments in ways that are largely alien to their training.

Still, as the techniques of digital humanities expand and become more commonly known and as they become more firmly integrated into institutions of higher education, with their own economy of prestige and reputation, digital humanists will be called on again to explain the aim of expanding the scope of factual knowledge about literature that can be collected and analyzed with its methods (to adapt an insight from Barbara Herrnstein Smith). In the process, digital humanities may need to distinguish its procedures from those Foucault

associates with the neoliberalism of behavioral economics. One answer might be that measuring literature in these ways is an intervention in the ongoing contest over what counts as data and how data becomes evidence in the analysis of culture. In some sense, the appeal to measurement and quantification by literary scholars may be a response, decades later, to the assertions of scholars like Becker that economic analysis can sufficiently explain the production of all cultural phenomena, including literature. Rather than see literature as economics in another form, literary criticism's use of measurable, quantifiable data offers a rejoinder to behavioral economics by asserting that literature possesses its own arithmetic, its own data that can be analyzed using tools adapted to its uniqueness.

Further Reading

References to Thomas Piketty are to *Capital in the Twenty-first Century* translated by Arthur Goldhammer (Cambridge, 2014). References to the French edition are to *Le capital au XXI^e siècle* (Paris, 2013).

David Harvey's review of Piketty is "[Afterthoughts on Piketty's Capital.](#)" For Larry Summers's comments, see "The Inequality Puzzle" in *The Democracy Journal* (Summer 2014): 92. For Paul Krugman's comments and review of Piketty's *Capital*, see "[Why We're in a New Gilded Age.](#)" *New York Review of Books* (May 8, 2014).

Stephen Marche's assessment of Piketty as a new kind of literary criticism can be read at "[The Literature of the Second Gilded Age.](#)" *Los Angeles Review of Books* (June 16, 2014).

Piketty made his remarks about Balzac's obsession with his debt originating his interest in the economic realities of wealth accumulation and inequality in an [interview](#) hosted by BBC's *Business Daily*.

For representative recent works related to behavioral economics, see Dan Ariely's *Predictably Irrational: The Hidden Forces that Shape our Decisions* (New York, 2010) and Tyler Cowen's *Discover Your Inner Economist: Use Incentives to Fall in Love, Survive your Next Meeting, and Motivate your Dentist* (New York, 2007) and *An Economist Gets Lunch: New Rules for Everyday Foodies* (New York, 2013).

Piketty is not the only economist who has recently appealed to literature and to Austen specifically to make arguments about economic models. See Branko Milanovic, *The Haves and the Have-Nots: A Brief and Idiosyncratic History of Global Inequality* (New York, 2010).

To read more about Gary Becker's ideas on marriage and how it relates to his other work on behavioral economics, see "A Theory of Marriage" in *The Economic Approach to Human Behavior* (Chicago, 1976).

Foucault's comments about Becker and neoliberalism are transcribed in *The Birth of Biopolitics: Lectures at the Collège de France, 1978-1979*, edited by Michel

Senellart and translated by Graham Burchell (New York, 2008).

Becker's reply to Foucault, and other thoughts on behavioral economics, were made in 2012. A recording of the interview and its transcript can be found at ["Becker on Ewald on Foucault on Becker"](#) interview, May 9, 2012.

For literary critics' response to Piketty's mode of literary analysis, see Ted Underwood, Hoyt Long, and Richard Jean So, ["Cents and Sensibility"](#) (2014) in *Slate*.

For a description of the process by which Underwood, Long, and So reached their conclusions, see Ted Underwood's talk delivered at Stanford Feb. 13, 2015, ["Piketty's Model: Literary History without Fixed Objects."](#)

Barbara Herrnstein Smith wonders about the aim of factual data in digital humanities in her talk "Scientizing the Humanities: Shifts, Collisions, Negotiations," forthcoming in *Common Knowledge*.

This article originally appeared in issue 16.3 (Summer, 2016).

James Mulholland is an associate professor of English at North Carolina State University and the author of *Sounding Imperial: Poetic Voice and the Politics of Empire, 1730-1820* (2013).