2021

On Not Already Knowing

J.D. Porter
University of Pennsylvania, jdporter@stanford.edu

ON NOT ALREADY KNOWING
J. D. Porter

A review of *Distant Horizons: Digital Evidence and Literary Change* by Ted Underwood. Chicago: University of Chicago Press, 2019. 200 pp. $83.00 hardback, $27.50 cloth.

Distant Horizons collects about a decade of cutting-edge literary digital humanities (DH) work into a concise, accessible volume. The five chapters work equally well as standalone experiments or in the service of Ted Underwood's overarching argument that DH reconfigures our understanding of literary history. It's a clear must read for anyone working in literary DH (especially text mining) or in literary history more generally, and when it is inevitably added to dozens of syllabi, both students and teachers will find a lot to admire. Yet the most profound achievement of the book is its demonstration of a genuinely new kind of literary critical knowledge.

This is not quite how Underwood casts it, though. He focuses on two methodological interventions made possible by the DH approach. First, there is the familiar issue of scale—the capacity of DH to tackle thousands of texts spread across centuries of production. Underwood argues that this new scale of attention fundamentally changes our understanding of literary history, writing that "we have narrated literary history as a sequence of discrete movements and periods because chunks of that size are about as much of the past as a single person could remember and discuss at one time" (ix). The digital approach, he says, enables us to consider changes that are too long term, slow moving, or widely dispersed to have been visible

to traditional methods. Chapter 1, for instance, shows that fiction grew increasingly distinct from nonfiction (especially biography and autobiography) over the course of 1800–2000, mostly on the basis of a rise in concrete language. Underwood approaches the problem from multiple angles, always with approachable, clear technical explanations, and arrives at the conclusion that literary history contains a large-scale pattern of change that has so far gone unnoticed by scholars. Yet the use of a long timescale may not entirely capture Underwood's achievement. It is not clear that people really do struggle to think about patterns of subtle and complex change on a 200-year scale (quick: Was a dollar worth more in 1818 or 2018?), and literary histories in particular often extend far past that—think of Erich Auerbach's *Mimesis* (1946), which stretches from Homer to Woolf. DH handles scale well and often, but it has forebears in traditional methods.

The second major intervention Underwood emphasizes is modeling, which is a hot topic in the DH world. To him, a model "defines a relationship between variables" as a way to study those relationships rather than "isolated facts" (ix). He views modeling, especially predictive and perspectival models, as a way to connect social and textual evidence, to mediate between theories and measurements. In chapter 2, Underwood demonstrates the historical stability of genre categories by showing how well certain models predict genre membership for novels in different periods, and by using models trained in one period to examine novels from another. The proof is in the pudding—in this as in every chapter, the results are both persuasive and interesting, as when Underwood shows that detective fiction has been remarkably cohesive as a genre since (but not before) Poe's Dupin stories. But again, Underwood may actually be underselling how radical his methods are. The recent discussion of models has never quite explained what *isn't* a model. Take *Mimesis* again: Isn't Auerbach's distinction between Hebrew and Greek literature in essence a model of historical literary thought, a way of relating variables (textual, social, theoretical) in order to provide an explanatory apparatus for complex literary data? Don't most critical approaches do that on some level?

What distinguishes the new empirical models is not that they are models, but that they are empirical. This is implicit in Underwood's emphasis on the rigorous comparison that his models enable, where rigor stands in for the many detailed and repeatable operations that numeric data allow—we cannot add and subtract the Hebrew or Greek approach to interiority, but we can do that and more with something we have counted. Underwood is very attentive to the affordances of numbers,

emphasizing their capacity to produce comparative assessments while maintaining an explicit distance from the naive positivism often attributed to DH scholars by skeptics of the field. He notes for instance that "numbers are not inherently more or less objective than words" (xviii) and argues that "the point of quantification can be to render description relative rather than objective" (67).

This comparative capacity produces a different kind of dizzying scale, a mind-boggling number not of years but of dimensions. Underwood's arguments typically depend on changes to thousands of individual words moving at varying rates across hundreds of volumes. The small and the large merge to produce something incomprehensibly complex—incomprehensible in the literal sense that a human being cannot picture a ball moving in five dimensional physical space, much less thousands of novels arranging themselves across a 3,000-dimension genre space. Underwood's models reduce this kind of information to something humanists—or really, humans—can use, dots that are numerous but all on the same plane, two or three hundred years arrayed along a simple line. All literary objects are complex, but *empirical* models enable a new way to engage with that complexity, to track the transition from text to model in a reproducible, reconfigurable, and,

crucially, reductive way. The many variables are latent in the novels and poems; Underwood's dimension reduction makes them visible.

The result is a new kind of knowledge. One of the most refreshing features of *Distant Horizons* is Underwood's recurring insistence that "the fact that something is retrospectively plausible doesn't mean we already knew it" (14). Given how frequently DH scholars still encounter the claim that "we already knew" whatever they just struggled to find, this seems destined to be one of the most widely cited elements of what is sure to be a widely cited book. And Underwood is supported by the opening up of two new kinds of knowledge, which we might think of in terms of the traditional philosophical definition of knowledge as justified true belief.

The first and easiest to see is a new form of justification. In the course of reaching chapter 4's finding that the language attached to male and female novel characters grew less distinct from about 1850 to 2000, Underwood also finds that the particular words shifted around—for example, "room" grows more and then less distinctly feminine, "grinned" more and less masculine. As he notes, "one possible conclusion would be that the structural positions of masculine and feminine identity, vis-á-vis each other, have remained very stable—while the actual content

of masculinity and femininity has been entirely mutable" (140). This is a classic case of something that we "already knew"; we can see versions of the same point everywhere from Judith Butler to the recent attention to the historical transition of computer programming from a job popularly associated with women to one associated with men. But now we have a new kind of evidence, the empirical changes in language across a large number of texts and time. Our beliefs are now justified in new ways, strengthening them for people who already held them and perhaps persuading new people.

Not all of Underwood's findings have such clear antecedents in famous critical works, however. It is often the case that, as he puts it, "[W]e barely have intuitions about patterns on this scale; our expectations are not clearly formed yet" (32). When he traces the history of prestige in the nineteenth and twentieth centuries, he finds that a model can accurately predict whether novels and poems were reviewed in journals solely on the basis of the words they use. Did we "already know" that prestige had so much to do with the distributions of a few thousand words? The idea may seem plausible, but surely most scholars never put the question to themselves in the first place. Not only can we now confirm things we had only guessed; we can assert things we had never guessed. This is the radical promise of *Distant Horizons*: The multidimensional empirical models of DH should create not just new justifications, but new beliefs—new ways to know.

J. D. Porter is a Digital Humanities Specialist in the Price Lab at the University of Pennsylvania. He wrote the pamphlet Popularity/Prestige *(2018), and his other work has recently appeared (or will soon appear) in* Episteme, Cultural Analytics, *and the anthology* Ralph Ellison in Context.