

INTELLECTUAL HISTORY AND DIGITAL HUMANITIES

DAN EDELSTEIN

Department of French and Italian, Stanford University

E-mail: danedels@stanford.edu

Peter de Bolla, *The Architecture of Concepts: The Historical Formation of Human Rights* (New York: Fordham University Press, 2013)

Jo Guldi and David Armitage, *The History Manifesto* (Cambridge: Cambridge University Press, 2014)

The digital age has been a boon for intellectual historians, particularly those of us who work on early modern Europe and America. The mass digitization of old books has made research more efficient than ever: first editions are there for the downloading on Google Books, Gallica, Liberty Fund, Project Gutenberg, and elsewhere. The creation of such large-scale databases as Early English Books Online (EEBO), Eighteenth-Century Collection Online (ECCO), the Making of the Modern World (formerly Goldsmiths’–Kress), or, on a more modest level, the ARTFL project’s FRANTEXT, has also breathed new life into old texts. Books that lay forgotten for generations can now be rediscovered thanks to the magic of search engines. To be sure, this power has not always been wielded for good: students today can “cite anything, but construe nothing,”¹ stringing together KWICs (keywords in context), and reading only a surrounding sentence or two (if that). But however they are used, these tools and platforms have transformed our daily work habits.

The same cannot be said, however, about our methods. Indeed, early modern intellectual historians still follow approaches that were established long before the Internet came of age. The two that continue to set the terms of our debates are the Cambridge school, whose major theoretical claims were put forward by Quentin Skinner in the 1960s; and the *Begriffsgeschichte* school, whose main tenets were established by Reinhart Koselleck in the 1970s. (A third important

¹ Jonathan Barnes, quoted by Anthony Grafton, “Codex in Crisis: The Book Dematerializes,” in *Worlds Made by Words: Scholarship and Community in the Modern West* (Cambridge, MA, 2009), 288–324, at 322.

tradition, associated with Warburg Institute scholars, still has many eminent practitioners—Anthony Grafton in chief—but tends to be less self-conscious about its ecumenical methods, which, in any event, are also rooted in the early twentieth century.) There are some exceptions to this rule: a handful of intellectual historians have ventured into the more sophisticated terrain of topic analysis, machine learning, word collocation analysis, and sequence alignment.² But where, say, literary critics now have multiple statements about how digital humanities can and should transform traditional methods of literary analysis, there has been nothing comparable in the field of intellectual history.³ (Spatial history, of course, is a different story.⁴)

Stepping into this void, but from very different directions, come the authors of the two books under review. Peter de Bolla has written an ambitious, promising, and somewhat frustrating book. *The Architecture of Concepts* is certainly not for the faint-hearted: de Bolla employs an idiosyncratic terminology that can turn prose into dense thickets of jargon, and his argumentation is largely proleptic, demanding great patience from the reader. But he engages in a valuable experiment that brings digital methods to the fore of intellectual history. Jo Guldi and David Armitage, for their part, have issued a rousing battle-cry to historians that is less demonstrative in its methodological prescriptions, but more compelling in its vision of where history should be headed. And while most

² See, respectively, David Newman and Sharon Block, “Probabilistic Topic Decomposition of an Eighteenth Century Newspaper,” *Journal of the American Society for Information Science and Technology*, 57/5 (2006), 753–67; Russell Horton, Robert Morrissey, Mark Olsen, Glenn Roe, and Robert Voyer, “Mining Eighteenth Century Ontologies: Machine Learning and Knowledge Classification in the *Encyclopédie*,” *Digital Humanities Quarterly*, 3/2 (2009), at <http://digitalhumanities.org/dhq/vol/3/2/000044/000044.html>; Keith Baker, “Revolution 1.0,” *Journal of Modern European History*, 11 (2013), 187–219; and Dan Edelstein, Robert Morrissey, and Glenn Roe, “To Quote or Not to Quote: Citation Strategies in the *Encyclopédie*,” *Journal of the History of Ideas*, 74/2 (2013), 213–36. Linguists have, of course, been working on similar questions, often using similar methods to the ones under discussion here: see, for instance, Derry Tanti Wijaya and Reyyan Yeniterzi, “Understanding Semantic Change of Words over Centuries,” in *Proceedings of the 20th ACM Conference on Information and Knowledge Management, workshop on DETecting and Exploiting Cultural diversiTy on the Social Web* (DETECT 2011), 35–40, at <http://dl.acm.org/citation.cfm?id=2064475>. My thanks to Melvin Wevers for this reference.

³ Examples from literary studies would include Franco Moretti, *Distant Reading* (New York, 2013); and Matthew Jockers, *Macroanalysis: Digital Methods and Literary History* (Urbana, IL, 2013).

⁴ See, most notably, Amy Hillier and Anne Kelly Knowles, eds., *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship* (New York, 2008).

historians are still groping their way around digital tools, the authors place digital humanities at the very heart of their recommendations.

Digital tools feature prominently in de Bolla's study, though their place is initially underplayed by the author. The chapter in which he describes his new approach to conceptual history is written without any mention of the digital humanities (the phrase does not occur in the book) or its foremost practitioners (Franco Moretti, for instance, goes unmentioned). There are a few sentences gesturing toward "the digital archive" and "powerful technology" (14), but de Bolla frames his methodological argument in terms of philosophy and cognitive science. Despite this odd disconnect between theory and practice, the two subsequently appear closely related in his work. And so, to appreciate the novelty of the practice, we must first consider the theory.

De Bolla is interested in how words map onto concepts, and how these maps can shift, sometimes imperceptibly, over time. His prime example of this movement, which also provides the case study for the book, is the difference between "rights of man(kind)" and "human rights." But he is also interested in the *unavailability* of certain concepts at different historical times. Could Enlightened writers even have thought "human rights" the way we do today? To advance such an argument, de Bolla needs to drive a wedge between words and the concepts they express: perhaps Locke granted all men natural rights, but what did he really mean by "rights" (or by "men")? In this regard, de Bolla's argument mirrors Foucault's famous claim about homosexuality; but where Foucault suggested that certain concepts were unthinkable without words, de Bolla reverses the relation to declare words insufficient to think certain concepts. Conceptual history thus becomes a method for determining when, and possibly why, words came to encompass different concepts.

The phrase "conceptual history" invariably brings to mind its German equivalent, *Begriffsgeschichte*. De Bolla is well aware of this historical school, though he does not situate his own method in relation to theirs, only citing Koselleck indirectly (if favorably). This is a pity, since his approach is ultimately quite similar: like Koselleck, de Bolla is interested in "contested concepts," and how rival definitions of the same word battle it out in public discourse over time.⁵ He prefers to draw a comparison with the Cambridge school, as personified by

⁵ It was in such terms that Koselleck described the methodological assumptions underpinning the *Geschichtliche Grundbegriffe*: "basic concepts are highly complex; they are always both controversial and contested." See Reinhart Koselleck, "A Response to Comments on the *Geschichtliche Grundbegriffe*," trans. Melvin Richter and Sally E. Robertson, in Hartmut Lehmann and Melvin Richter, eds., *The Meaning of Historical Terms and Concepts: New Studies on Begriffsgeschichte* (Washington, DC, 1996), 59–70, at 64. See more generally Melvin Richter, "Koselleck on the Contestability of 'Grundbegriffe':

Quentin Skinner. De Bolla emphasizes similarities before homing in on one key difference: where Skinner seeks to identify “the mental landscape of an agent in the past,” he wishes “to recreate the cultural terrain that provided historical actors with a conceptual lexicon for playing out their roles” (30). In other words, where Skinner is concerned with what a given author meant in a given text, de Bolla wants to know what an array of authors could mean in an array of texts at a given time. More than a “conceptual turn” in intellectual history (4), what de Bolla is really proposing is more like a cultural turn.

As the cultural turn was a few turns back, it may seem that de Bolla is spinning in circles. But his proposal does point in new directions when translated into the world of databases and search engines, which allow researchers to explore the aggregate in far more exhaustive ways than before. The answer to the famous question “what do you do with a million books?” posed by digital humanities pioneer Gregory Crane, is obviously not to read each one to determine its meaning. Where quantification allows us to scale up from the one book to the many, however, it only displaces, rather than resolves, the problem of figuring out what the numbers mean.

Finding meaning in large-scale, aggregated results is the challenge that de Bolla sets himself, using the ECCO database as his case study. As he acknowledges, it is a far from ideal resource: digitized by optical character recognition (OCR), the texts, in their digital incarnations, are frighteningly inaccurate.⁶ The database also contains multiple editions of the same works, making frequency counts difficult to interpret. Finally, the search interface is quite restricted, though de Bolla makes the most of it, using search operators to conduct collocation analyses for terms within n words of each other. With this method, he produces a series of tables listing the most common terms that appear in the vicinity of “rights,” breaking down the results in twenty-year spans. He uses these data to track the rise and fall of related concepts over the course of the eighteenth century: for instance, at the beginning of the century (1700–1720), “divine” appeared within five words of “rights” slightly more often than “man,” whereas by the end of the century (1780–1800), the latter term would be found ten times more commonly in its vicinity (seventy-nine). Similar stories can be told for the terms “humanity,” “equal,” “constitution,” “property,” “nature,” and “people” (all of which occur with increasing frequency near “rights”), as well as for “church,” “ancient,” “prerogative,” “royal,” and “majesty” (in relative terms, decreasing frequency).

A Comparative Perspective,” in Carsten Dutt and Reinhard Laube, eds., *Zwischen Sprache und Geschichte: Zum Werk Reinhart Kosellecks* (Gottingen, 2013), 69–95.

⁶ For an even more damning account than de Bolla’s see Patrick Spedding, “‘The New Machine’: Discovering the Limits of ECCO,” *Eighteenth-Century Studies*, 44/4 (2011), 437–53.

De Bolla admonishes his reader to focus on his method rather than his precise results, and for good reason. First, he produced his collocation tables by hand: they are not an automated list of the most commonly found terms within *n* words of one another, but rather a list of those terms de Bolla noticed occurring most frequently. As far as I could tell, the list is fairly accurate (one important term that is not included is “bill”), but his method would be painful to reproduce.⁷ Second, his approach raises questions about data normalization. There are 2.5 times as many documents in ECCO for the period between 1780 and 1800 as for 1700–1720 (71,052 versus 28,944). De Bolla recognizes that raw numerical results can be misleading, but argues that the numbers are not final: they are there to suggest general patterns and to solicit interpretations (9).

With all his qualifiers in place, the results that de Bolla has obtained are insightful and thought-provoking. The real question is, how much can we read into them? De Bolla suggests a number of interpretations, some more convincing than others. He begins by noting how the term “duties” is used increasingly within five words of “rights,” an observation he uses to challenge (albeit gingerly) Richard Tuck’s thesis about the allegedly subjective quality of rights onward from the seventeenth century (66–73). He then makes a case for a “conceptual grammar” of rights, taking as his example the terms “rights”/“liberties”/“privileges” (91–101). This section is less convincing, since de Bolla overlooks the fact that “rights, liberties, and privileges” (or variants thereon) was a stock phrase in English political discourse, already in wide circulation in fifteenth-century texts (as searching the EEBO database attests).

A great deal of de Bolla’s book is concerned with the emergence of what he claims was “a new concept, the rights of man,” which he distinguishes from “the rights of *men*.” He suggests that this former concept first appeared, only to quickly vanish, in the run-up to the American Revolution, before gaining widespread currency in England, after 1780. He devotes an entire chapter to the impact of Thomas Paine’s *Rights of Man* on this terminological and, in his view, conceptual development, but fails to consider a more obvious cause for the success of this expression: it translates the French expression *droits de l’homme*, made famous by the 1789 Declaration of that name. Indeed, of the 4,157 documents published between 1780 and 1800 that feature the words “rights” and “man” within five words of each other, 3,460 (or 83 percent) were published on or after 1789; of these, 2,958 (or 85 percent) also include the word “France.” We can find

⁷ The ARTFL project runs a smaller, more accurate version of the ECCO database (ECCO-TCP) on its PhiloLogic search and retrieval engine, which automatically generates collocation tables. See <http://artfl-project.uchicago.edu/content/ecco-tcp>. I was thus able to compare the list of terms identified by PhiloLogic within five words of “rights” with de Bolla’s lists.



Fig. 1. (Colour online) Frequency rates for “rights of man,” “rights of men,” “rights and privileges.” Source: Google Ngram Viewer, at <https://books.google.com/ngrams>.

an even clearer indication of this French connection by searching for the exact phrase: “rights of man” can be found in 436 documents published between 1700 and 1788, and in 2,975 documents published between 1789 and 1800 (inclusively, for both date ranges); normalizing for total number of documents, the phrase appears in twenty-three times more documents in the final decade of eighteenth century. Rather than the result of gradual conceptual shifts in English discourse, or the even the publication of Paine’s influential book, the sudden explosion of this term is more likely explained by events across the Channel.

This remark leads to a more general problem with de Bolla’s method. While he writes about changes in *frequency* in word collocation, this is not technically correct: his searches only return the number of *documents* containing the proximate terms, not the frequency with which the terms co-occur. In other words, his results place on the same plane texts that mention “rights” near “man” once in five hundred pages, and those that cite them together on every page. The ECCO search interface does not provide the functionality needed to obtain true frequency results, but a casual experiment with Google’s Ngram Viewer reveals how this method would paint a very different a picture (Fig. 1).

Figure 1 shows just how great the spike “rights of man” was in the immediate aftermath of the French Revolution, but how this was when the expression “rights of men” took off as well. This coincidence in real frequency makes de Bolla’s attempt to tease these expressions apart conceptually less convincing—no doubt both expressions were being used in large part to translate the French *droits de l’homme*. Examining real word frequencies also suggests that it was actually in the 1760s (and not 1780–1800, as de Bolla argues) that “rights and privileges” was employed most commonly (a similar pattern holds for “rights and liberties”). Of course, Google Books contains different material than ECCO, so one would need to confirm these results on the same sources.

What is the promise of de Bolla's approach for intellectual history more generally? To his credit, de Bolla does not push his quantitative findings farther than they can comfortably go. He recognizes that the numbers alone are open to different interpretations, but can suggest promising avenues of exploration. At the same time, his "architectural" method is best suited for tracking this large-scale ebb and flow of conceptual relationships over time. It starts to break down when he comes to the reading of individual texts. This is unfortunately the case with the book's central argument, the distinction between between "rights of *man*," which he claims "could never be held by men" (142), and the "rights of *men*" which could be enumerated and defined. But to defend this claim, de Bolla is often reduced to speculating about the presence of one or the other concept in a given text. Often his evidence is extremely thin: his claim that the American delegates hit upon "a new concept, the rights of man," in October 1774 rests on a single letter; what's more, this letter employs the commonplace expression "natural rights of mankind."

De Bolla made the most of the search functionality afforded by a commercial database. But these technical limitations can and should be overcome, and must not be seen as fundamental restrictions on the promise of text analysis. Most research libraries hold the actual text files for ECCO; while awkward to manipulate, access to these files allows users to perform a much broader variety of sophisticated analyses. In de Bolla's case, topic modeling and real frequency analysis would allow him to refine and revise his results, and also offer a better grasp on the granular level. This is not to suggest that digital tools should always replace the very valuable analogue techniques of reading and interpretation; oftentimes, these tools are most helpful for indicating which parts of a large corpus we ought to read more carefully. But having a broader range of tools and methods is important for textual analysis, as it allows the researcher to slice across a corpus in different ways, ideally to cross-validate results, but also to reveal irregularities.

As their title suggests, Jo Guldi and David Armitage's book is a big, bold, visionary statement about the future of historical studies, and, even more importantly, about the "public mission" (123) that they argue once fell to historians, and has since been lost, but can be recovered in the future. Their focus is not specifically on intellectual history, though their arguments about historical methodology are pertinent to this field as well—with, perhaps, a few qualifiers.

Two concepts, one old, one new, underpin *The History Manifesto*. The first is Fernand Braudel's famous theory of the *longue durée*, a methodological axiom of the *Annales* school. The other is the much more recent notion of "big data."

Guldi and Armitage argue that by combining these two concepts—exploring lots of data over long stretches of time—historians can craft bigger and better narratives that will capture the attention of policymakers, as well as the public.

The *Annales* historians also worked with large datasets, but what distinguishes contemporary data practices is in fact a third element: data visualization. Guldi in particular is well placed to assess the merits of visualization, being one of the creators of Paper Machines, an open-source visualization platform that runs on the Zotero reference management system. As the authors observe, the advances in visualization techniques and user experience that computers have facilitated now means that historians can download, refine, and visualize streams of data on a scale and at a speed unimaginable during the heyday of the *Annales*, not to mention just ten years ago. What's more, an increasing number of these visualization platforms are designed for and by historians.⁸ Guldi and Armitage can thus imagine a near future in which “scholars will be able to take on a much larger body of texts than they normally do” (91).

This is no futuristic pronouncement, as evidenced by the examples in *The History Manifesto* from Guldi's forthcoming study *The Long Land War*, which is grounded in her use of Paper Machines to analyze the fluctuating prominence of different colonies (later countries) in British debates about land reform. She uses as her corpus “large numbers of bureaucratic texts on global land reform from the twentieth century” (92), which in turn points to one of the origins of the big data that Guldi and Armitage celebrate: “The arrival in the past ten years of mass digitisation projects in libraries and crowd-sourced oral histories online announced an age of easy access to a tremendous amount of archival material” (93). But the authors are not content with historians occupying the role of mere consumers in this data “revolution” (a term which occurs repeatedly in their manifesto). They also believe that historians should roll up their sleeves and start designing their own tools: “Historians may become tool-builders and tool-reviewers as well as tool-consumers and tool-teachers . . . If historians . . . take up this challenge, they may find themselves in the avant garde of information design” (114). This might sound like wishful thinking, but again, Guldi's own experience codesigning Paper Machines shows it is no mere bravado. What's more, if historians do not actively involve themselves in the design of digital research tools, they will be forced to rely on tools that were often designed with nonhistorical datasets in mind, along with nonhistorical research questions. A case in point would be Gephi, which some historians have used with great success, but whose force-directed graphs can impose a false sense of proximity

⁸ Disclosure: I am a principal investigator for one of the visualization projects they discuss (“Mapping the Republic of Letters”).

and distance when one is working with an incomplete dataset (such as a correspondence network) rather than, say, a data dump from Twitter.⁹

With its open-access publication (a first for Cambridge University Press), and concluding summons (“Historians of the world, unite!”), *The History Manifesto* is meant to inspire. It tells a riveting story of how economists replaced historians as policy experts in international and national institutions, and also of how historians have narrowed the chronological focus of their research over the past century. For intellectual historians, the practical question is one of determining how relevant these methodological claims are to our field. Armitage had already made the case that “big history” could and should apply to intellectual history as well, but had not rested this earlier claim on the availability of big data.¹⁰

To the extent that intellectual historians work with texts, Guldi’s example, along with de Bolla’s book, clearly shows that the practices of intellectual history are compatible with big data. But there are more kinds of data than texts. Libraries store vast holdings of precious metadata, currently locked up in pre-Internet-era MARC records, which make any kind of sophisticated querying nearly impossible. The current move toward “linked data,” however, will unlock these data for downloading, manipulation, and visualization, largely to the benefit of intellectual historians.¹¹ Other data troves are already up and running, such as the Early Modern Letters Online union catalogue of correspondence metadata, hosted at Oxford University.¹²

If we are well poised as intellectual historians to make use of large data sets in our research, the quality and completeness of these data are another issue. As we saw, ECCO contains terribly inaccurate texts; but, perhaps even graver, the extant historical record in many places (particularly correspondence) can be very patchy. In this respect, many of the data with which intellectual historians traditionally work do not have the same quality as the “untapped sources of historical data” (96) that the authors celebrate (including “data about democracy, health, wealth, and ecology” (100). While we can and no doubt should work more with the data available to us, this does not necessarily mean that we can or should form a “new school of quantitative analysis” (97). Visualizing data can be extraordinarily helpful for identifying trends, noticing holes, and setting the backdrop for an

⁹ For some successful uses of Gephi by historians see Emma Rothschild, “Isolation and Economic Life in Eighteenth-Century France,” *American Historical Review*, 119/4 (2014), 1055–82; and the Six Degrees of Francis Bacon project at <http://sixdegreesoffrancisbacon.com>.

¹⁰ See David Armitage, “What’s the Big Idea? Intellectual History and the *Longue Durée*,” *History of European Ideas*, 38/4 (2012), 493–507.

¹¹ For an example of what this future will resemble, readers can explore the open-data portal of the French National Library at <http://data.bnf.fr>.

¹² See <http://emlo.bodleian.ox.ac.uk>.

argument, but that is still a far cry from cliometrics. Not only must we recognize the limits of what our data can tell us (in terms of their exhaustivity), but we must also continue to cultivate the skills of interpretation. Rarely do numbers alone tell the full story. Historians should not have to become economists to replace them.

This is a worry that the authors address, arguing that historians are in fact better suited than social scientists to conduct *longue durée* studies of big data, because we are already trained to be more skeptical and open about our sources. Indeed, they even suggest that “the arbitration of data” sources will become “a role in which the History departments of major research universities will almost certainly take a lead” (107). Again, this is an ambitious project, particularly if we are also to be designing and building tools (and presumably studying a little history on the side). While very laudable and in some respects necessary, Guldi and Armitage’s proposals do raise the question of opportunity costs—particularly as concerns about time-to-degree continue to mount. More practically, *when* is the budding historian to master all these skills? In college? Graduate school? As an assistant professor? After tenure? These questions are not meant to challenge Guldi and Armitage’s agenda, but rather to air more mundane concerns about its implementation.

In the end, the authors recognize that our greatest strength as historians is one we already have: the power to tell good stories. They stress this point in conclusion, emphasizing the “need for new narratives capable of being read, understood, and engaged by non-experts” (117). And they also return here to the importance of data visualization, but no longer simply as a method of identifying trends in large data sets. Just as important, they argue, is its use as a rhetorical trope for conveying information in an elegant, convincing fashion (“We also need informative visualisations of our research and to put them in public” (119)). One reason we lost out to the economists was because they have better charts. But we are also better storytellers, and should use this skill to “parse the data of anthropologists, evolutionary biologists, neuroscientists, historians of trade, historical economists, and historical geographers, weaving them into larger narratives that contextualise and make legible their claims and the foundations upon which they rest” (112). As might be expected from the title, *The History Manifesto* is a daring pitch to make history *la reine des facultés*. At least the throne is currently vacant.