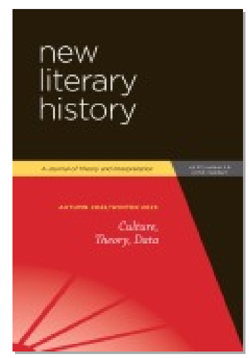
Situated Knowledges and Partial Perspectives: A Framework for Radical Objectivity in Computational Social Science and Computational Humanities

Laura K. Nelson



New Literary History, Volume 53, Number 4, Autumn 2022 / Volume 54, Number 1, Winter 2023, pp. 853-877 (Article)

Published by Johns Hopkins University Press
DOI: *https://doi.org/10.1353/nlh.2022.a898331*

Situated Knowledges and Partial Perspectives: A Framework for Radical Objectivity in Computational Social Science and Computational Humanities

Laura K. Nelson

> Just as the invention of the telescope revolutionized the study of the heavens, so too by rendering the unmeasurable measurable, the technological revolution in mobile, Web, and Internet communications has the potential to revolutionize our understanding of ourselves and how we interact. . . . Three hundred years after Alexander Pope argued that the proper study of mankind should lie not in the heavens but in ourselves, we have finally found our telescope. Let the revolution begin.
>
> —Duncan Watts

We have entered a new era of scholarship in the social sciences and humanities: the computational era. Rooted in the analysis of unprecedented amounts of social data and cultural material, computational social science and computational humanities blend traditions from the social sciences and the humanities with tools and approaches from computer science, engineering, and linguistics. This disciplinary cross-pollination has unearthed old disciplinary debates, including those around epistemology: what constitutes knowledge and the goals of knowledge building and, particularly in the interpretive social sciences and the humanities, the role of objectivity in knowledge creation.

Objectivity in science and scholarship has always been a contested terrain; it is no different in the computational era. While there are complexities and nuances to debates about objectivity, broadly speaking since the nineteenth century,[1] epistemological debates in science have revolved around the appropriate role of the subject in generating objective knowledge.[2] On one side of this discourse is the ideal of science and scholarship as the pursuit of disembodied objectivity.

This god-like view from nowhere objectivity—objectivity defined as the absence of subject—seeks to render the world as it is, independent of any observer. On the other side is the pursuit of embodied objectivity. Embracing subjectivity as constituent of objectivity, embodied objectivity emerges via the interplay between knowing subject and the situated object of knowledge.

For those who embrace the ideal of disembodied objectivity, the computational era is seen as the culmination of, or at least the next step in, the quest for a subjectless (social) science. In this ideal, the intervention of the subject into the process of knowledge creation is needed only when phenomena are not able to be represented empirically. With ever increasing access to data, we can now simply render the world as it is, without the need for subjective theory, interpretation, or methods.[3] For others, the computational era has instead reinforced the necessary role of the subject in producing any form of knowledge, from contextualizing the object of knowledge to interpreting output.[4] The disembodied approach to computational social science and humanities has thus far been the dominant approach, as evidenced by highly cited and well-publicized articles and books in the most impactful academic outlets (e.g., *Science*, *Nature*, *PNAS*, and prestigious university presses). But others contend that the embodied approach to objectivity is much more aligned with the foundational structure of computational methods.

This essay starts from the premise that the systematic production of knowledge (what some call science) is a worthy—even utopian—pursuit and that scholars in the social sciences and humanities should strive toward the creation of objective knowledge about ourselves, humanity, and society. From this premise, I argue four things: (1) Applying new digital data and computational tools available to us toward the goal of disembodied objectivity profoundly misunderstands the nature of the methods and undermines the revolutionary potential of the current moment for the social sciences and humanities. (2) Instead, large volumes of digitized cultural material and the computational methods to analyze that material are perfectly aligned with the embodied vision of objectivity, and these tools enhance our ability to achieve this vision of objectivity in ways not previously possible. (3) Compared to the disembodied objectivity perspective, embodied objectivity enables us to better leverage computational tools to produce more accurate accounts of the world, long sought by practitioners of these methods. (4) Borrowing from feminist theorists' take on embodied objectivity, I propose a version of Donna Haraway's articulation of situated knowledges and partial perspectives (SKPP) as an alternative framework for the application of computational methods in the social sciences and humanities, one that can facilitate the revolutionary potential of this new era of scholarship.

Objectivity and Its Discontents

Objectivity has never had a single definition. Its meaning is historically situated and, within each historical period, its meaning is hotly contested.[5] Perhaps the most common meaning of objectivity in the modern and contemporary era is its disembodied version. The path of science since the nineteenth century has, by and large, been one of progressively moving the knowing subject farther and farther away from the object of knowledge, through scientific tools (e.g., the microscope, telescope, and surveys) and procedures (e.g., statistics and experiments), culminating (for some) in the robots and the artificial intelligence of the twenty-first century.[6]

While this vision of objectivity is still dominant in the sciences, it is far from universally accepted. Allan Megill proposed four principal senses of objectivity present in epistemological discussions: *absolute objectivity*, *procedural objectivity*, *disciplinary objectivity*, and *dialectical objectivity*.[7] The first three of these senses share the perspective that objectivity requires eliminating or controlling the subject. Absolute objectivity is the "view from nowhere" objectivity that seeks to eliminate, quite literally, the subject from the creation of knowledge. Procedural objectivity, or the use of systematic methods or rules (such as statistics and experiments), begrudgingly acknowledges the necessary role of the knowing subject but seeks to limit it to one who carries out impersonal standardized procedures. Disciplinary objectivity, or objectivity determined via the consensus of a particular research community, aims to contain subjectivity via scientific debate and discussion, with a particular focus on the role of communication in producing objective knowledge. What Megill calls dialectical objectivity (and what I call here embodied objectivity) is alone in embracing subjectivity as necessary for producing objective knowledge.

Debates around the role of subjectivity in knowledge production reached a local peak during the mid-twentieth-century structuralist trend that swept through many disciplines across the social sciences and humanities. Feminist standpoint theory emerged in response. Grounding her work in structuralism itself, Marxist theorist Nancy Hartsock argued that the different structural location of men and women, particularly vis-a-vis the relationship of men and women to the means of reproduction, affords women a different epistemological vantage point in understanding the structure of society compared to men.[8] Standpoint theory developed out of this foundational observation. Personal experiences and perspectives are not only unavoidable in science and scholarship, but, because of distinct epistemological vantage points, are actually necessary to produce the most accurate and encompassing accounts of the world.[9] Transforming the structural locations of different groups of people into both sites of inquiry and sources of insights necessitates expanding what counts as data and what counts as objects of knowledge. Women's perspectives on the mundane, everyday processes that structure the social world, for example, can be transformed into sites of inquiry via institutional ethnography.[10] Songs, poetry, and oral traditions capture perspectives and knowledge from those traditionally excluded from scholarly production.[11] The canon does not capture the entirety of literature worthy of analysis.[12]

By providing theories about the relationship between social location and epistemology, as well as new methods and sources of data to transform social location into measurable sites of inquiry, standpoint theorists provided a scientific framework that maintained the goal of generating better accounts of the world, but it was one that refused to participate in the ideology of disembodied objectivity. Knowledge can not be considered objective, standpoint theorists ultimately argued, if its most basic truths—that all views are necessarily views from somewhere, that all objects of knowledge are situated somewhere—are denied.[13]

The new computational era has collided with this clash over objectivity in both predictable and novel ways.

Reviving Social Physics

The most visible response to the computational era has been from those who are enthralled with the potential of these methods to enhance disembodied objectivity. Much as photography was embraced by scientists in the nineteenth century for its potential to accurately and directly capture the world as it is, massive amounts of data have been embraced for their potential to reflect directly the state of the world in any given moment.[14] The digitization of books, for example, was celebrated for its potential to report "raw" facts about literary history objectively, including the rise and fall of genres, the structure of narratives, and trends in the gender breakdown of authors.[15] And digitized books, many believed, could do far more than quantify literary trends. Google Books, a corpus of digitized text containing about 4% of all books ever printed, promised to be able to accurately quantify cultural trends in "fields as diverse as lexicography, the evolution of grammar, collective memory, the adoption of technology, the pursuit of fame, censorship, and historical epidemiology."[16] Coining the term *culturomics*, those who constructed the Google Books corpus believed that this approach would expand "rigorous quantitative inquiry" (read: disembodied objectivity) to cultural fields.[17] This particular version of the macroanalysis of literature and culture is important not simply for its incorporation of many more books as objects of knowledge but also because, according to its early practitioners, the reporting of raw facts is more objective, and thus more scientific, precisely because their presentation does not require any interpretation on the part of the knowing subject.[18]

The expansion of quantitative methods into more so-called "subjective" scientific fields is echoed in another response to the computational era: reinvigorating the concept of social physics. August Comte's original idea for *sociology* was a science of society that could be "considered in the same light as astronomical, physical, chemical, and physiological phenomena," one that would seek out the "natural and invariable laws" that govern ourselves, humanity, and society, and one that he imagined would be the crowning achievement of all the sciences.[19] Network science and complexity studies have embraced the social physics framework and updated it for the contemporary computational era. In his book *Social Physics*, for example, Alex Pentland proposes that our habits of action are wired into us through our coordination in social groups, seemingly by instinct and without cognition.[20] Similarly, Mark Buchanan, a scientist trained in physics, claims that people can be viewed as "atoms or molecules following fairly simple rules" and that computational methods, in their capacity to analyze social data to infer those rules, are the "'quantum revolution' in the social sciences."[21]

Highly visible publications aligned with this vision of computational methods are vast. Preferential attachment, also known as the Matthew Effect or the rich-get-richer phenomenon, is one of the more prominent examples of a supposed simple rule that governs social interactions, and evidence for preferential attachment (e.g., power law distributions) have been identified in data as diverse as biological, physical, and social networks.[22] In cultural fields, Matsumae et al. used complex distance measures and network analysis on genetic, lexical, grammatical, and musical data to identify underlying laws that, they argue, can explain the historical development of both the genetics and culture of populations.[23] Scholars used *Seshat*, a global history database compiled to be the most comprehensive body of knowledge about human history, to claim that the presence of moralizing gods are necessary for the development of complex societies.[24] And so on.

Identifying these simple rules structuring societies is only one approach in this social physics tradition. Others contend that it may not be the case that societies follow Newtonian-type laws but rather that collectives exhibit predictable statistical patterns, much like food chains or ant or bee colonies.[25] Individual events or actions, such as avalanches, individual emotions, and particular life courses, are difficult to predict, but all avalanches, emotional trends, and general life courses have common statistical patterns.[26] Individuals, in other words, may be variable, but their combined action "lead[s] to predictable outcomes for the collective," and the goal of computational analyses should be to identify these common statistical patterns.[27]

This view motivates computational scholars to search for universal patterns in large amounts of diverse data. Scott Golder and Michael Macy, for example, used sentiment analysis on Twitter data from multiple continents to show that "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures."[28] Satyam Mukherjee et al. used data from the Web of Science and identified "the nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology": a common temporal pattern in citations among hit science papers across multiple fields and over time.[29] Lu Liu et al. used the same data and found that individual success across fields as diverse as science, law, and culture tends to happen in hot streaks.[30]

Others in this tradition go beyond describing statistical patterns to making predictions using data. Google Flu Trends, for example, one of the early "successes" of "big data," aggregated Google search activity in order to predict flu outbreaks.[31] More recently, Princeton sociologist Matt Salganik hoped to demonstrate the power of these new methods to predict life courses through a prediction competition for social science using data from the Fragile Families and Child Wellbeing Study, one of the largest and most detailed representative collections of longitudinal social data.[32]

The initial excitement over the potential of these methods was followed by a predictable trough. While there may be broad patterns in collectives, practitioners soon acknowledged that societal changes and specific social patterns may be triggered by a few key individuals, individuals with agency and intent. This "law of the few," and the fact that there is so much fluctuation around the averages for individual behavior, makes it difficult to scale from individual behavior—the social data that form the base of the data used in social physics—to groups or collectives.[33] Social physicists were dismayed that "universal" patterns tended to not be so universal in social systems. It is not unusual, found science writer Philip Ball, "for one model or theory to appear to be supported by experience in one situation, but others in other circumstances."[34] Ball concluded that, rather than searching for universal models that hold for all groups and societies, social physics (or complexity studies) should instead embrace "pluralistic modeling" or "several complementary and overlapping models, some of which work well sometimes or for some aspects of a problem and others in other cases."[35]

The track record of the early empirical studies in this tradition has been rocky at best. Moretti's and Jockers's early analyses of literary trends, for example, were based on data that were incomplete, and, at times, wrongly labeled.[36] Because Google Books is not a systematic sample of any collection of books or cultural material, scholars have argued that trends identified using the corpus may simply be a reflection of what Google has digitized, not of any real cultural or linguistic trends.[37] Challenges identified by historians in the way the Seshat data were coded were serious enough to warrant a retraction of the original article on moralizing gods.[38] In 2013, Google Flu Trends missed the peak of the 2013 flu season by a whopping 140%, prompting the shutdown of the service and examinations into why it failed.[39] A large study of about one thousand different networks found that Barabási's "universal" scale free networks are actually relatively rare, occurring in only around 4% of the networks studied.[40] When Salganik and team tested the models submitted to their Fragile Families challenge, they were surprised: "Despite using a rich dataset and applying machine-learning methods optimized for prediction," they found that even the most sophisticated machine learning models and the ones that performed the best "were not very accurate and were only slightly better than those from a simple benchmark [regression] model."[41] He and his team concluded that their exercise suggested "practical limits to the predictability of life outcomes in some settings."[42] Duncan Watts, one of the early leaders in the field of computational social science, has reached a similar conclusion through his decades of research. The real world is complex, he claims, and history only plays out once. "It's not that we can't predict anything," concludes Watts; "the problem is that the predictions we most want to make are precisely the ones we can't make."[43]

While early attempts at using any new methods and data are almost always improved upon, these ongoing substantive problems suggest a more fundamental issue for computational social scientists and humanists. When these methods are assumed to eliminate the knowing subject from the knowledge production process, when they are pursued with the goal of disembodied objectivity, scholars are motivated to hide the many subjective decisions required to implement any computational project and as such, they dangerously overstate the accuracy and reliability of their results. Instead, as data gets larger, and complex computational and algorithmic processes remove scholars farther away from the raw data, the role and perspective of the knowing subject becomes more, not less, important. Subjective decisions start from the very first step of computational projects: choosing which data to use and claiming what processes that data capture.[44] Once the data are decided, additional subjective decisions are required throughout computational projects: choosing what to search for in large, complex data requires theory, knowledge, and intuition; which features of the data are used dramatically impacts the resulting representation of the data; choosing an algorithm and parameters alters results; and results on their own are meaningless without some form of interpretation. Because of the sensitivity of computational methods to the content of the data—they are designed to explicitly learn patterns directly from data—each of these subjective decisions directly impacts the patterns identified and conclusions drawn. Raw facts, it turns out, are not so raw.

The reality of massive data and computational methods is undeniable: despite enthusiasm from social physicists and cultural empiricists, the way computational methods are designed and implemented aligns the computational era much more with embodied objectivity than with disembodied objectivity.[45]

Enhancing Embodied Objectivity

Computational methods and digitized data enable the generation of embodied knowledge along multiple dimensions and in ways feminist and other standpoint theorists have long sought. First, the amount and diversity of data available means we can now include more and more diverse viewpoints as objects of knowledge. Feminist theorists have been critical of mainstream social science and humanities scholars for their at times inordinate focus on data and cultural material produced by, or capturing the perspective of, (white) men. Contemporary digital trace data captures a multitude of social and cultural processes, including the mundane processes and relational discourse that feminist theorists have long sought to directly capture. Expanded access to historical material, including books from the margins, pamphlets and magazines, advertisements, social pages, and obituaries in newspapers, capture diverse sources of historical knowledge from groups traditionally excluded from science and scholarship. While there are, of course, still gaps and omissions in available material, the amount and diversity of contemporary and historical data we now have access to expands the perspective from which we can analyze material and diversifies what counts as phenomena that can be empirically measured, studied, and communicated.[46]

Second, the richness and complexity of this digitized material captures the situated and embodied nature of the world it represents. Digital trace data do not force complex categories such as race and gender into columns on a spreadsheet, abstracted from their cultural and historical contexts. The data used in computational projects are much richer in depth and context, including text and images that convey detailed perspectives, relational data that capture complex social networks, and longitudinal and geographic data that allow for temporal and geographic comparisons, situating perspectives in social, cultural and historical contexts.[47]

Third, computational methods, in particular machine learning, model the depth and complexity of the multiple perspectives captured in large data while preserving cultural and social associations embodied within. Computer scientists were initially dismayed that their supposedly subjectless algorithms were absorbing social and cultural biases contained in and conveyed through cultural material.[48] Social scientists and humanists instead saw enormous analytic promise in the ability for these algorithms to extract these associations embedded in language and images—what Ted Underwood calls perspectival modeling.[49] Machine learning methods are revolutionary for the social sciences and humanities precisely because they model[50] and mimic[51] the way humans learn and communicate cultural associations—the very processes standpoint theorists have argued should be the basis of knowledge production.

Fourth, far from removing the subjectivities of the researcher from the analytic process, computational methods demand more subjectivity from the researcher at every step of the process. Embodied objectivity calls for embracing the interplay between subject and object and documenting the role of various subjectivities in the generation of knowledge. Computational methods encode (literally, in the case of programming languages) each of these subjective decisions that go into the analysis process, making visible those decisions, subjecting them to open discussion and debate, and, more broadly, making the role of the subject in knowledge production more transparent.[52]

The following two empirical examples, both situated at the intersection of the social sciences and humanities, illustrate how computational methods can be used in the service of embodied objectivity.

Example 1: Online Publics and Counterpublics

Twitter has exploded over the past decade as an object of research in both the social sciences and the humanities, due in part to the public nature of Twitter, the role of Twitter in early digitally enabled social movements such as the Arab Spring and Occupy Wall Street, the networked nature of Twitter—with hashtags, mentions, and retweets providing natural links between nodes to model—and because of the relative ease of collecting large amounts of Twitter data. Twitter was one of the initial inspirations behind the revival of the social physics perspective, but it has also become central to critical scholarship.[53]

A paper by computational social scientist Brooke Foucault Welles and scholar Sarah Jackson illustrates the potential to capture multiple embodied perspectives in large amounts of data, modeling the ways in which different structural locations interact with the same online platform, leading to distinct uses of that technology.[54] Jackson and Foucault Welles sought to analyze the consolidation of national discourse around police violence in the US following the 2014 fatal shooting of Michael Brown by a police officer in Ferguson, Missouri—one of the events that catapulted the Black Lives Matter movement onto the national scene and an early example of the role of counterpublics in shaping Twitter discourse. Following Brown's shooting and death, local activists used #Ferguson on Twitter to share developing news about the shooting and to frame Brown's death in terms of broader trends of police violence against Black men. Picked up by high-profile activists and journalists, the hashtag quickly went national, raising the profile of the antipolice violence and Black Lives Matter movements.

Like many early Twitter studies, Jackson and Foucault Welles sought to understand how the #Ferguson discourse developed and spread online. Following best practices used in Twitter research, they scraped tweets that included #Ferguson and used social network analysis on retweets and mentions in tweets to identify the most influential Twitter users tweeting about Brown's death—what they called crowdsourced elites. Their initial analysis identified high profile activists and journalists as central crowdsourced elites, but it completely missed local activists that one of the authors knew were crucial to the early moments of this discourse. Knowing their analysis missed important actors shaping this discourse, they redid their analysis, creating a separate network model for each day, starting with the day Brown was shot.

By identifying which users were most central day by day rather than in the aggregate, they found that the early crowdsourced elites were local Black activists, mostly Black women, and the discourse was primarily shaped by one local Black woman, @AyoMissDarkSkin. Her tweets framed Brown as an innocent victim of extreme violence by the police—a framing that was then propagated throughout Twitter and formed the basis of the critical interpretation of Brown's death. Alternative and niche media, previously found to have been central drivers of digital storytelling around conflicts between citizens and the state, were, they found, almost entirely absent from the first week of #Ferguson discourse, as were politicians and entertainment figures. As #Ferguson started to trend nationally, alternative media, politicians, and entertainment figures with large followings moved to the center of the network, both amplifying and changing the original framing. Because these later central actors had such large followings, they dominated the data when it was analyzed in the aggregate—what most network scholars continue to do.

Jackson and Foucault Welles's research highlights three important features of computational research when it is done from an embodied rather than disembodied perspective. First, Jackson and Foucault Welles used their own perspectives, in particular their qualitative and critical knowledge of the way marginalized groups use Twitter, to question and interrogate their methods as they were implementing them. By allowing their own subjective knowledge to shape their analytic choices, they arrived at an arguably more accurate account of discursive dynamics on Twitter than they would have had they simply followed the accepted method. Second, they used the richness of digital trace data to identify and model multiple perspectives in the same data, situating each of these perspectives in the context of the lives of those generating the data. Local Black activists, whose social position meant they had privileged access to the way Brown's killing impacted the local community but who also did not have large Twitter followings, used Twitter in a particular way: to share information with their community and to craft a common critical message around Brown's killing to advance a political agenda. National actors, who were in a much different structural position vis-a-vis Brown's killing and their positions on Twitter, used Twitter to report, comment on, and amplify political trends happening on the local level. Local activists used this feature of Twitter—the ability to be retweeted by those with large followings—to compel the mainstream to pay attention to their cause and to consciously shift narratives about Black communities. Jackson and Foucault Welles were able to capture this complex dynamic by questioning the initial "universal" statistical patterns their method identified, instead (re)modeling the data to capture multiple perspectives. Third, their analysis exemplifies the risk of aggregating large amounts of data: statistical aggregation silences intersecting and marginal perspectives in data. The outcome of this statistical silencing is not purely a social justice issue; it also leads to inaccurate (and thus not objective) accounts of the social world. These authors concluded that in order to use computational methods to accurately model the social world and to ensure methods do not marginalize important perspectives in large amounts of data, scholars must "make big data small again" by leveraging its richness rather than its size.

Example 2: Measuring Intersectionality

Intersectionality is an ideal conceptual instantiation of the embodied perspective on objectivity. The concept of intersectionality initially drew attention to the false universality of concepts such as gender and race, emphasizing instead that these categories are experienced differently across social groups. Machine learning, I argue, is ideally suited to measure, or give a visible form to, intersectionality, a contention I illustrate here via an analysis of intersectional experiences of the nineteenth-century US South using word embeddings.[55]

Intersectionality as a concept is used across disciplines and thus has a variety of definitions.[56] For my purposes, I define intersectionality as a theoretical framework for understanding how social identities and categories combine with each other and interact with systems of social, cultural, economic, and political power to create distinct, and unequal, lived experiences. Word embeddings are a machine learning technique that takes a corpus as input and outputs a high-dimensional vector space model of the corpus. A vector is an object that contains components (typically numbers) that represent data within a set space (for example, X, Y coordinates on a two-dimensional plot). Word vectors are simply sets of numbers that represent the meaning of the word based on the context in which the word appears across a corpus. The word vectors can then be compared to one another spatially. In theory, word vectors that are closer to each other within the vector space (typically measured via cosine similarity) are semantically similar to one another. Scholars have found that in addition to mapping linguistic similarities, word-embedding models can reveal shared cultural stereotypes embedded in language.[57] When used on first-person narratives, I argue, word embeddings can spatially map differences in the lived experiences[58] across intersecting identities, perfectly capturing the definition of intersectionality proposed above.

To use word embeddings to capture intersectional experiences, I compiled a corpus from *Documenting the American South*, a collection of first-person narratives from the US South published during the long nineteenth century. The editors of this collection purposefully sought to capture diverse experiences during this century, including from enslaved and formerly enslaved people, Black and white women, soldiers, and laborers. The diversity of narratives included, as well as the fact they were collected to directly convey the experiences of people with distinct perspectives on this era, made this an ideal corpus to measure intersectional experiences of this century. After some cleaning of the collection (for example, removing multiple editions of the same book), my corpus for this analysis included 414 book-length, first-person narratives from two of their subcollections: "First-Person Narratives of the American South" and "North American Slave Narratives."[59]

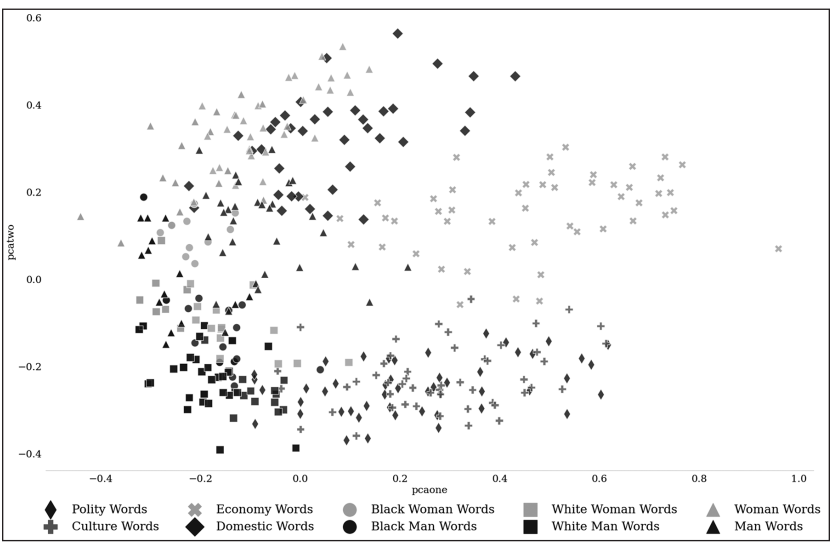
I matched my analytic approach as close as possible to the definition of intersectionality offered above. After training a word-embedding model on this corpus, I used two techniques to model the relationship between people's intersecting identities in the context of different social and cultural institutions to measure how the interrelationships between identity and institutional context are implicated in lived experiences. To represent intersecting identities I used vector addition, averaging across synonym pairs that approximate social identities.[60] I constructed vectors for Black and white men and women by creating a list of words for women (*woman, girl, girls, she, her, hers, herself*), men (*men, man, boy, boys, he, him, his, himself*), Black (*black, colored, coloured, negro, negress, negros, afroamerican*), and white (*white, caucasian, anglosaxon*). I then summed the vectors resulting from the vector addition over every possible combination of pair-wise categories (e.g., *women* + *black*) and divided the resultant by the number of total possible pairs for each pair-wise category combination. I took the average of the synonyms for *women* and for *men* to construct generic women and men vectors.

To capture the different institutional contexts in which social identities are embedded, I constructed four separate vector spaces (extracted from the larger word embedding space) to represent four institutions that shape how different forms of power are distributed in society: the polity (political power), the economy (economic power), culture (cultural power), and the domestic (capturing differences in participation in the public and private spheres). To identify these subspaces, I identified the fifty vectors with the highest cosine similarity to *nation* + *state* (representing the polity), *money* (representing the economy), *culture,* and *housework* + *children* (representing the domestic). To this institutional subspace, I added the fifty vectors with the highest cosine similarity to two metaidentities (men and women) and four intersecting social identities (Black and white men and women). This resulted in a 476-by-476 matrix, with rows and columns as the 500 words with the highest cosine similarity to at least one of the four social institutions or six social identities (less the words that were similar to more than one of these categories), and cells as the cosine similarity between each of the pairs of words. In theory, this high-dimensional space spatially represents cultural associations between intersecting identities and social institutions as they appear in this corpus.

As it is impossible to visualize high-dimensional space directly, I first used principal component analysis (PCA) to render this space in two dimensions. Figure 1 visualizes the first two components of the resulting

PCA. Interpreting the spatial locations of the words in this visualization, the first principal component (*pcaone*) captures individual words (the left side of the graph, where most of the social identity words are clustered) compared to collective words (the right side, where most of the social institutional words are clustered). This dimension is not surprising or particularly informative, as the way I designed this subspace included a combination of individual social categories plus collective social institutions. In a sense, this dimension provides face validity to this way of visualizing these data and this space. The second dimension (*pcatwo*) is more informative and, in my interpretation, captures aspirational compared to practical discourse. Individuals and nations must meet their practical needs: individuals need to eat and have shelter, and nations must provide the structures to enable people to meet these needs. If their practical needs are met, both individuals and nations can then have aspirational desires: culture and refinement in living conditions and in character, democratic governance and ideals of nation-states, and so on. In Figure 1, practical discourse is captured on the top of the

Fig. 1. First two dimensions from a PCA of 476 words with highest cosine similarity to each of four institutions and six social categories

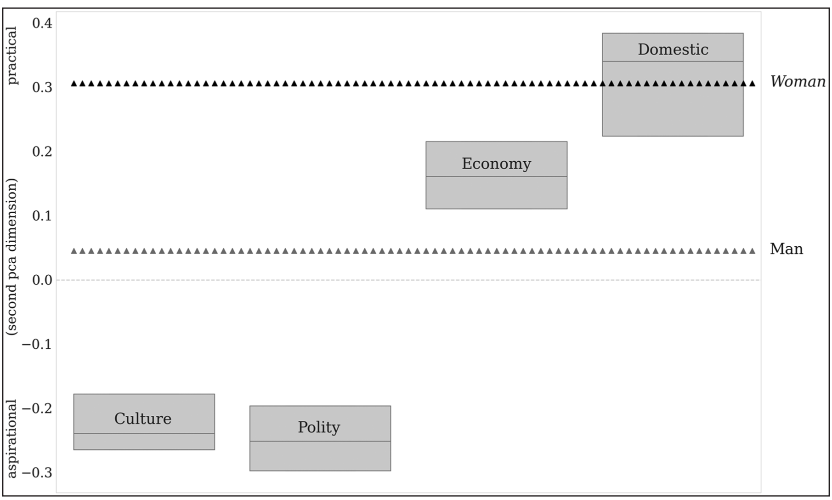


The figure visualizes the first two dimensions from a principal component analysis of the fifty words closest to six social category vectors (Black and white men and women, plus women and men) and four social institutions. I interpret the first dimension, pcaone, as distinguishing between individual (left) and collective (right) words. I interpret the second dimension as distinguishing between practical (top) and aspirational (bottom) words.

Y-axis, clustered largely around domestic and economic words, while aspirational discourse is on the bottom, driven largely by political and cultural words.

Visually, it appears that words closer to the *woman* and *Black woman* vectors might be clustered at the top of this second dimension, while those closer to *man*, *white woman*, and *white man* might be clustered at the bottom of the dimension. This practical/aspirational dimension thus may be one way to understand intersectional experiences captured in this corpus. Figure 2 visualizes how social identity is related to this second component (aspirational compared to practical). The gray box plots in Figure 2 represent the distribution of words for each of the four social institutions across pcatwo. Culture words and polity words are on average clustered on the aspirational pole of this dimension, while economy and domestic words are on average clustered on the practical pole of this dimension. The lines in Figure 2 represent the average distance of the fifty words associated with two social identities—*woman* and *man*—to pcatwo. The *woman* vector is located highest along this dimension,

Fig. 2. Distribution of words closest to four social institutions along pcatwo and the average distance from pcatwo for two social categories



The y-axis is the second dimension from a principal component analysis of the fifty words closest to six social category vectors and four social institutions. The gray box plots represent the distribution of fifty words with the highest cosine similarity to each of four social institutions across the second pca dimension. The lines represent the average distance to the second pca dimension of the fifty words with the highest cosine similarity to two social identities.

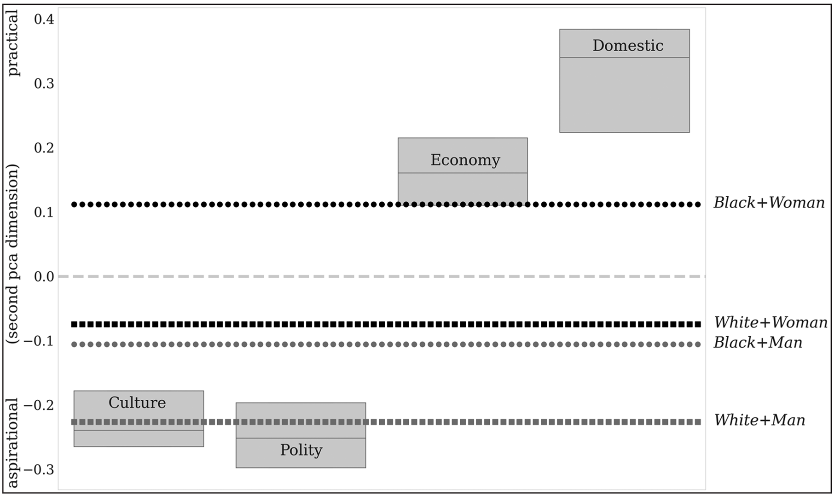
similar to the distribution of the domestic words, while the average for the the *man* vector is lower on this axis, falling in the middle of the polity, culture, and the economy distributions (see Figure 2). Figure 2 thus almost perfectly captures classic theories of gender, with women confined to the private sphere and, in turn, to practical concerns.

If I was following the social physics model and was searching for universal patterns in large amounts of data, my next step might be to repeat this analysis on different corpora, or gather a much larger corpus (such as Google Books) and test whether a similar gender schema is universally present. The theory of intersectionality suggests an alternative analytic path: Black women will have had a very different lived experience of the nineteenth-century US South compared to white women and Black men. These data and methods allow us to visualize these differences. To measure intersectional experiences of this century, I repeated the above steps but for the *black+woman*, *black+man*, *white+woman*, and *white+man* vectors, visualized in Figure 3.

Figure 3 confirms that the position of *man* and *woman* along the practical/aspirational axis is only a partial perspective of gender schema in this corpus. While both the *white+woman* and *black+woman* vectors were closer to the practical pole of this axis compared to *white+man* and *black+man*, supporting the gender theory described above, there are also important within-gender differences that a gender universalist approach misses. The average position of the fifty words similar to the *white+woman* vector is much lower on the Y-axis, closer to the aspirational pole. This is largely due to the proximity of the *white+woman* vector to culture words. The words similar to the *black+woman* vector are instead much higher on the Y-axis, closer to the practical pole, and are clustered predominantly around the economy words.

The words with the highest similarity to the *culture* vector in this corpus are words such as *endowments*, *refinement*, *thrift*, *acquirement*, and *intellectual*: words that describe the perceived high moral and cultural superiority of southern *gentility* (another word similar to the *culture* vector) as conveyed in these first-person narratives. Because of nineteenth-century US racial schemas, these descriptors were reserved only for white people (the *white+men* vector is equally proximate to, and *black+men*, similarly far away from the culture words). In this corpus, and as illustrated in Figure 3, while white women were celebrated for their endowments, refinement, and gentility (aspirational concerns), Black women were celebrated for their ability to make money, their industry, and their thrift (practical concerns).

Fig. 3. Distribution of words closest to four social institutions along pcatwo and the average distance from pcatwo for four intersecting social categories



The Y-axis is the second dimension from a principal component analysis of the fifty words closest to six social category vectors and four social institutions. The gray box plots represent the distribution of fifty words with the highest cosine similarity to each of four social institutions across the second pca dimension. The lines represent the average distance to the second pca dimension of the fifty words with the highest cosine similarity to four compound social identities.

By spatially locating social-identity vectors alongside four social institutions, I used word embeddings to visualize the perspective of different social locations embedded within different institutions. The result is an empirical account of the precise ways in which gender and race intersected in this context to produce differing proximities to various social and cultural spheres. In addition, this analysis inductively identified the aspirational/practical axis as one major axis organizing gender and racial schemas in this corpus. While women on the whole were associated with practical concerns, as gender schema would predict, proximity to cultural descriptors (gentile, refined, etc.) gave white women access to the aspirational sphere in ways Black women were not allowed, and Black women were associated with economic descriptors in ways white women were not. Black men's similar proximity to economic markers and distance from cultural descriptors associated them with the practical sphere in ways that white men were not.

This is, of course, only a partial perspective on the question of intersectionality during this century. This is not a random sample of literature from this century; it is an expertly curated corpus that leans largely proabolitionist. A different corpus—one that leans proslavery—would provide a different perspective on this same century. Furthermore, other social categories, such as class, and other institutions, such as religion and authority, could provide even more perspectives on this same question. A benefit of this approach is that others can implement this method from different perspectives on the same data or on data from different historical contexts, providing a way to systematically compare intersectionality across identities, institutions, contexts, and temporal periods.

The Situated Knowledges and Partial Perspectives Framework

The above two examples illustrate the differences between an embodied objectivity application of computational methods and the disembodied objectivity approach. I propose four principles to guide this embodied approach to computational social sciences and humanities:

P1: The object of knowledge, and thus objectivity itself, is embodied

The object of knowledge in social science and the humanities—social data and cultural material—is embodied. Each data point objectifies a historical and cultural moment. If data are inanimate (genres, periods, states, communities), each data point was constructed or constituted by humans and embodies the perspectives and biases of those creating/constituting it. If the data are about individuals, each data point has intent, goals, and motivations, as well as a gender, race, class, and, quite literally, a body with a weight, height, and an aesthetic. The embodied nature of a datum impacts what it represents, what information it can convey, and/or how the individual it represents moves through the world, how it interacts with others, and how it is reacted to by others. Any knowledge created from social and cultural data must acknowledge and specify how the embodied nature of their data impacts their conclusions and the knowledge generated from their conclusions. Any method or approach used should not abstract out of this embodied reality but should embrace it in order to construct more accurate representations of the culture and society the analyzed data and cultural material represents.

P2: Only partial perspectives promise objectivity

Every view is a view from somewhere, and this includes both the knowing subject and the object of knowledge. Objectivity is constructed by piecing together a series of partial perspectives, specifying from which perspective a particular knowledge claim holds. The difficult work of knowledge creation is not in making universal claims absent the knowing subject—something that cannot in fact ever be achieved—but in being precise and transparent about the contours of the partiality of any and every knowledge statement.

P3: Technology enables accurate, transparent, and reproducible rendering of partial perspectives

The key to objective partial perspectives and situated knowledges is the ability to measure the world through different perspectives—treating each perspective not as a point of view, but as a site of inquiry. Computational methods, and machine learning in particular, provide a useful tool to represent different perspectives in data and cultural material, allowing those perspectives to be sites of inquiry through which we can generate partial knowledge statements that are reproducible, transparent, and communicable to a research community.

P4: The object of knowledge is an actor and agent, not a resource

Knowledge created via computational social science/humanities has the potential to directly impact, in both negative and positive ways, the subjects generating that data. Ethical, fair, and transparent computational social science and humanities rests on the foundational acknowledgment that the object of knowledge in social science and the humanities is an actor or agent; it is not a resource to be exploited. Not only do partial perspectives and situated knowledges allow for radical objectivity in the social sciences and humanities, but because this framework acknowledges the embodiment and situatedness of data, this framework also encourages researchers to be answerable and responsible for what we learn how to see.

Generating Generalizable Knowledge

For some in the humanities and parts of the interpretive social sciences, the goal of scholarly work is to characterize and draw insights from particularities and the exceptional. When applied using the SKPP framework and in the service of embodied objectivity, computational methods can enhance this goal. Computational methods can be used, for example, to identify "islands of difference" within large amounts of diverse data: subgroups of data or cultural material that diverge from dominant patterns and may be worthy of more detailed engagement. Understanding how cases are particular or exceptional can then augment insights drawn from detailed engagement with each case.

Others in the humanities and social sciences seek instead to identify broader social and cultural patterns or to use data to *explain* social or cultural processes.[61] Generating generalizable knowledge and/or explanations using the SKPP framework requires some additional tools, borrowed from qualitative methodologies. In particular, qualitative scholars in the social sciences have spent decades theorizing how to build generalizable knowledge from case studies. While computational projects in the social sciences and humanities deal with large amounts of data, these projects are still dealing with a single or a few cases. In the social sciences, for example, Twitter is a single case, as are Reddit and Instagram. In the humanities, genres,[62] publishing companies,[63] or selected literary works from a particular time period or geographic location[64] are each examples of single (or a few) cases that can be used to generate knowledge that extend beyond each particular case. In the examples provided above, the #Ferguson discourse was a single case, as was the corpus of first-person narratives of the US South. Computational social scientists and humanists interested in building more general knowledge should ground their computational projects in one or more theories of case studies.

Single cases, for example, can be used to construct what sociologist Josh Pacewicz calls constitutive arguments: context-independent analytical descriptions that describe categories in the social or cultural world.[65] Constitutive arguments, constructed using a single or a few exceptional or more typical cases, can be used to reveal the limits of existing or standard social or cultural categories, they can point to and empirically describe new classes of objects or social processes not currently covered in existing categories, and/or they may be used to magnify relational patterns that drive social and cultural processes—relational processes that may not be visible in other circumstances or cases.[66] Cases, in other words, can be used to push the boundaries or challenge existing ways of categorizing and explaining the world. Alternatively (or additionally), the point of comparison can be causal mechanisms, not categories: cases can be used to identify potential mechanisms that can explain causal connections between social or cultural processes.[67]

Case studies can also be extended temporally or geographically, for example, to identify variations in the case across time or space.[68] As demonstrated in the two empirical examples above, these temporal or geographic variations, especially when they are surprising or challenge existing theory,[69] can then be used to reconstruct or refine existing theory, constructing more accurate and richer explanations of the world, extended case by extended case.

While building knowledge from case studies—through constitutive arguments, causal mechanisms, or reconstructing or refining social theory—computational social scientists and humanists should additionally be clear about their analytical lens: what is the researcher trying to explain? When implementing a project aimed at building more generalizable knowledge, scholars should specify (at least) three analytic lenses.[70] What is the *level* of explanation? Is it the micro (e.g., individuals, cultural artifacts, individual relationships), the meso (e.g., communities, genres, organizations), or the macro (e.g., states, periods, societies)? What is the *subject* of explanation? Is it people or artifacts, places (e.g., social media platforms, geographic locations) or mechanisms? What is the *location* of explanation? Is it durable dispositions—explaining why people behave or believe similarly regardless of the social situations, or why cultural patterns are replicated across time or space—or is it explaining how the social situation (e.g., Twitter) impacts how actors behave and believe, or how the cultural and historical context (e.g., eighteenth-century Britain) impact what patterns in cultural material mean?

Combining a theory of a case with analytical lenses can help push the SKPP framework beyond a patchwork of different perspectives, which are valuable in and of themselves, to generating knowledge that can help us describe and explain broader patterns in social and cultural worlds.

Conclusion

One reaction to the intellectual hubris exhibited by some using computational methods, and the real-world harm that has come from it, has been to double down on the idea that objectivity is a myth. This, I argue, is a mistake. In a world of alternative facts and a deluge of mis and disinformation, objectivity has once again become a tool of the oppressed. Objectivity is not only a laudable goal, but in this new computational era, we have the ability to pursue it in ways not previously possible, while embracing the embodied and reflexive epistemologies supported by many in the social sciences and humanities.

While objectivity is a worthy pursuit, believing that one data set, one algorithm, or one model can capture a universal, subjectless truth remains a dangerous myth. Situated knowledges and partial perspectives was initially proposed to redirect scientific and knowledge-building processes away from totalizing ideologies. This framework can do the same in this new computational era. Whether knowledge production proceeds via the particular and exceptional, or via using cases to build generalizable knowledge, the SKPP framework provides a path to constructing objective knowledge without recourse to the ideology of disembodied objectivity. This, I argue, is the next step for computational social science and humanities, one that realizes the objectivity that makes science a worthwhile pursuit. This is the revolution in knowledge that has long been the promise of this new era of scholarship.

University of British Columbia

NOTES

1 Lorraine Daston and Peter Galison, *Objectivity* (New York: Zone Books, 2007).
2 Andrew Gelman and Christian Hennig, "Beyond Subjective and Objective in Statistics," *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 180, no. 4 (2017): 967–1033.
3 Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired*, June 23, 2008, https://www.wired.com/2008/06/pb-theory/; Matthew Lee Jockers, *Macroanalysis: Digital Methods and Literary History* (Urbana: Univ. of Illinois Press, 2013); Monica Lee and John Levi Martin, "Coding, Counting and Cultural Cartography," *American Journal of Cultural Sociology* 3, no. 1 (2015): 1–33; and Franco Moretti, *Graphs, Maps, Trees: Abstract Models for Literary History* (London: Verso, 2007).
4 Katherine Bode, "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History," *Modern Language Quarterly* 78, no. 1 (2017): 77–106; Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (Cambridge, MA: MIT Press, 2020); Gelman and Hennig, "Beyond Subjective and Objective in Statistics"; Richard Jean So and Edwin Roland, "Race and Distant Reading," *PMLA* 135, no. 1 (2020): 59–73; Ted Underwood, "Machine Learning and Human Perspective," *PMLA* 135, no. 1 (2020): 92–109.
5 Daston and Galison, *Objectivity*.
6 Evelyn Fox Keller, "The Paradox of Scientific Subjectivity," in *Science and the Quest for Reality*, ed. Alfred I. Tauber (London: Palgrave Macmillan, 1997), 182–200.
7 Allan Megill, "Introduction: Four Senses of Objectivity," in *Rethinking Objectivity*, ed. Megill (Durham, NC: Duke Univ. Press, 1999), 1–20.
8 Nancy C. M. Hartsock, "The Feminist Standpoint: Developing the Ground for a Specifically Feminist Historical Materialism," in *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science*, ed. Sandra Harding and Merrill B. Hintikka (Dordrecht: Springer Netherlands, 1983), 283–310.

9 Harding, *The Science Question in Feminism* (Ithaca, NY: Cornell Univ. Press, 1986).
10 Dorothy E. Smith, *The Everyday World as Problematic: A Feminist Sociology* (Boston: Northeastern Univ. Press, 1987).
11 Patricia Hill Collins, *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment* (New York: Routledge, 2009).
12 Deborah E. McDowell, "New Directions for Black Feminist Criticism," *Black American Literature Forum* 14, no. 4 (1980): 153; and Lillian S. Robinson, "Treason Our Text: Feminist Challenges to the Literary Canon," *Tulsa Studies in Women's Literature* 2, no. 1 (1983): 83.
13 Donna Haraway, "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective," *Feminist Studies* 14, no. 3 (1988): 590.
14 Daston and Galison, *Objectivity*.
15 Moretti, *Graphs, Maps, Trees*, esp. 14–33 and 53–55.
16 Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331, no. 6014 (2011): 176.
17 Michel et al., "Quantitative Analysis of Culture," 176.
18 Jockers, *Macroanalysis* is another example.
19 Georg G. Iggers, "Further Remarks about Early Uses of the Term 'Social Science,'" *Journal of the History of Ideas* 20, no. 3 (1959): 434.
20 Alex Pentland, *Social Physics: How Social Networks Can Make Us Smarter* (New York: Penguin Books, 2015).
21 Mark Buchanan, *The Social Atom: Why the Rich Get Richer, Cheaters Get Caught, and Your Neighbor Usually Looks like You* (New York: Bloomsbury USA, 2007), x.
22 Albert-László Barabási, "Scale-Free Networks: A Decade and Beyond," *Science* 325, no. 5939 (2009): 412–13.
23 Hiromi Matsumae et al., "Exploring Correlations in Genetic and Cultural Variation across Language Families in Northeast Asia," *Science Advances* 7, no. 34 (2021).
24 Harvey Whitehouse et al., "RETRACTED ARTICLE: Complex Societies Precede Moralizing Gods throughout World History," *Nature* 568, no. 7751 (2019): 226–29.
25 Philip Ball, *Why Society is a Complex Matter: Meeting Twenty-first Century Challenges with a New Kind of Science* (Berlin: Springer, 2012).
26 Ball, *Why Society is a Complex Matter*, ix.
27 Buchanan, *The Social Atom*, xxi.
28 Scott A. Golder and Michael W. Macy, "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures," *Science* 333, no. 6051 (2011): 1878–81.
29 Satyam Mukherjee et al., "The Nearly Universal Link between the Age of Past Knowledge and Tomorrow's Breakthroughs in Science and Technology: The Hotspot," *Science Advances* 3, no. 4 (2017).
30 Lu Liu et al., "Hot Streaks in Artistic, Cultural, and Scientific Careers," *Nature* 559, no. 7714 (2018): 396–99.
31 Jeremy Ginsberg et al., "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature* 457, no. 7232 (2009): 1012–14.
32 Matt Salganik, *Fragile Families Challenge*, *https://www.fragilefamilieschallenge.org/* (accessed September 4, 2021).
33 Guido Caldarelli, Sarah Wolf, and Yamir Moreno, "Physics of Humans, Physics for Society," *Nature Physics* 14, no. 9 (2018): 870.
34 Ball, *Why Society is a Complex Matter*, xii.
35 Ball, *Why Society is a Complex Matter*, xii.
36 Bode, "The Equivalence of 'Close' and 'Distant' Reading."
37 Eitan Adam Pechenick, Christopher M. Danforth, and Peter Sheridan Dodds, "Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution," *PLOS ONE* 10, no. 10 (2015).

38 Edward Slingerland et al., "Historians Respond to Whitehouse et al. (2019), 'Complex Societies Precede Moralizing Gods Throughout World History,'" *Journal of Cognitive Historiography* 5, no. 1–2 (2018): 124–41.
39 David Lazer et al., "The Parable of Google Flu: Traps in Big Data Analysis," *Science* 343, no. 6176 (2014): 1203–5.
40 Anna D. Broido and Aaron Clauset, "Scale-Free Networks Are Rare," *Nature Communications* 10, no. 1 (2019): 1017.
41 Matthew J. Salganik et al., "Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration," *Proceedings of the National Academy of Sciences* 117, no. 15 (2020): 8398.
42 Salganik et al., "Measuring the Predictability of Life Outcomes," 8398.
43 Duncan Watts, quoted in Marina Krakovsky, "The Anti-Predictor: A Chat with Mathematical Sociologist Duncan Watts," *Scientific American*, March 29, 2011, https://www.scientificamerican.com/article/duncan-watts-book/.
44 Bode, "The Equivalence of 'Close' and 'Distant' Reading"; and Juan Pablo Pardo-Guerra and Prithviraj Pahwa, "The Extended Computational Case Method: A Framework for Research Design," *Sociological Methods & Research* 51, no. 4 (2022): 1826–67.
45 D'Ignazio and Klein, *Data Feminism.*
46 Roopika Risam, *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy* (Evanston, IL: Northwestern Univ. Press, 2018).
47 Bode, "The Equivalence of 'Close' and 'Distant' Reading."
48 Tolga Bolukbasi et al., "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings," *ArXiv:1607.06520*, July 21, 2016, http://arxiv.org/abs/1607.06520.
49 Underwood, *Distant Horizons: Digital Evidence and Literary Change* (Chicago: Univ. of Chicago Press, 2019).
50 Austin C. Kozlowski, Matt Taddy, and James A. Evans, "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings," *American Sociological Review* 84, no. 5 (2019): 905–49.
51 Alina Arseniev-Koehler and Jacob G. Foster, "Machine Learning as a Model for Cultural Learning: Teaching an Algorithm What It Means to Be Fat," S*ociological Methods & Research* 51, no. 4 (2022): 1484-1539.
52 Laura K. Nelson, "To Measure Meaning in Big Data, Don't Give Me a Map, Give Me Transparency and Reproducibility," *Sociological Methodology* 49, no. 1 (2019): 139–43.
53 Deen Freelon, Charlton McIlwain, and Meredith Clark, "Quantifying the Power and Consequences of Social Media Protest," *New Media & Society* 20, no. 3 (2018): 990–1011; and Sarah J. Jackson, Moya Bailey, and Brooke Foucault Welles, *#HashtagActivism: Networks of Race and Gender Justice* (Cambridge, MA: MIT Press, 2020).
54 Jackson and Foucault Welles, "#Ferguson Is Everywhere: Initiators in Emerging Counterpublic Networks," *Information, Communication & Society* 19, no. 3 (2016): 397–418.
55 A replication repository with the code and data needed to reproduce the analysis in the section "Example 2: Measuring Intersectionality" can be found here: *https://github.com/lknelson/RadicalObjectivity.*
56 Patricia Hill Collins, "Intersectionality's Definitional Dilemmas," *Annual Review of Sociology* 41, no. 1 (2015): 1–20.
57 Kozlowski, Taddy, and Evans, "The Geometry of Culture"; and Austin van Loon and Jeremy Freese, "Word Embeddings Reveal How Fundamental Sentiments Structure Natural Language," *American Behavioral Scientist* 67, no. 2 (2022): 175–200.
58 I use the phrase "lived experience" to capture the social milieux that produces intersecting social identities that are contingent (i.e., not biologically determined) but experienced in ways that individuals do not choose and cannot alter. See Simone de

Beauvoir, *The Second Sex* (New York: Vintage, 2011).

59 "First-Person Narratives of the American South," Documenting the American South, accessed July 2, 2020, https://docsouth.unc.edu/fpn/; "North American Slave Narratives," Documenting the American South, accessed July 2, 2020, https://docsouth.unc.edu/neh/.

60 Following the method proposed by Kozlowski, Taddy, and Evans in "The Geometry of Culture."

61 For example, see Rens Bod, *A New History of the Humanities: The Search for Principles and Patterns from Antiquity to the Present* (Oxford: Oxford Univ. Press, 2013).

62 Rachael Scarborough King, "The Scale of Genre," *New Literary History* 52, no. 2 (2021): 261–84.

63 So, *Redlining Culture: A Data History of Racial Inequality and Postwar Fiction* (New York: Columbia Univ. Press, 2020).

64 Bode, "The Equivalence of 'Close' and 'Distant' Reading"; and Hoyt Long, *The Values in Numbers: Reading Japanese Literature in a Global Information Age* (New York: Columbia Univ. Press, 2021).

65 Josh Pacewicz, "What Can You Do With a Single Case? How to Think About Ethnographic Case Selection Like a Historical Sociologist," *Sociological Methods & Research* 51, no. 3 (2020): 931–62.

66 Ivan Ermakoff, "Exceptional Cases: Epistemic Contributions and Normative Expectations," *European Journal of Sociology* 55, no. 2 (2014): 223–43.

67 George Steinmetz, "Odious Comparisons: Incommensurability, the Case Study, and 'Small N's' in Sociology," *Sociological Theory* 22, no. 3 (2004): 371–400.

68 Michael Burawoy, "The Extended Case Method," *Sociological Theory* 16, no. 1 (1998): 4–33; Guerra and Pahwa, "The Extended Computational Case Method."

69 Iddo Tavory and Stefan Timmermans, *Abductive Analysis: Theorizing Qualitative Research* (Chicago: Univ. of Chicago Press, 2014).

70 Colin Jerolmack and Shamus Khan, "The Analytic Lenses of Ethnography," *Socius: Sociological Research for a Dynamic World* 3 (2017): 1-11.