A HETEROGLOSSIA OF BIG DATA KNOWLEDGES

Thylstrup, Agostinho, Ring, D'Ignazio, and Veel's *Uncertain Archives: Critical Keywords for Big Data* is in many ways an extraordinary piece of scientific work. Consisting of 640 pages with 61 chapters authored by no less than 73 contributors, the book offers a substantial addition to existing scholarship on the societal consequences of the large scale datafication of modern life. Yet it is not only the sheer size of the book or its well-crafted composition that is remarkable, but also—and more importantly—the rich insights it presents.

The general tone is critical, and the chapters share a concern for big data archives and their promises of "previously unknown certainty." The editors argue that along with the high hopes for big data come a "whole host of uncertainties equally unknown to humankind" (p. 1). In the era of big data, the notion of the archive has moved from a "regime of knowledge about the past to a regime of future anticipation," with big tech companies telling us that they now have command of everything from cultural trends to criminal acts, epidemics to environmental disasters, and terrorist threats. Recent information scandals related to electoral fraud, illegal surveillance, and biased predictive policing systems have caused researchers and observers not only to question the statistical validity of data-driven analyses, the editors argue, but also to consider broader societal implications of big data's determination of knowledge. Such implications are the focus of the book.

The voices included are diverse: while some authors draw on resources from established disciplines such as sociology, philosophy, and computer science, others respond with interdisciplinary methodologies and experimental approaches. There is a predominance of female authors, yet identities across a broad spectrum of gender and geography are represented. As

a result, the collection provides multiple perspectives on, and insights into, the aesthetic, cultural, political, and technological aspects of big data and algorithms. Even the encyclopedia-mimicking format of the book emphasises this point: different lines of arguments intersect and interconnect, with no dominating or unifying narrative leading the way. In my opinion, this is both one of the book's greatest strengths and one of its potential weaknesses.

As hinted in the title, the book is organised in 61 chapters—the 61 "critical keywords"—presented in alphabetical order. While many of these terms will be familiar to readers interested in questions of information and data science, others are less obvious. Entries such as "Aggregation" (Sune Lehmann), "Database" (Tahani Nadim), and "Outlier" (Catherine D'Ignazio) may appear more recognisable than, say, "Figura" (Frederik Tygstrup), "Flesh" (Romi Ron Morrison), or "Throbber" (Kristoffer Ørum). It is refreshing to see such a breadth of topics: a breadth that is mirrored in the different writing styles and perspectives provided by a polyphony of voices. This "multivocal and dialogical" nature of the book, the editors argue, can be seen as a performative enactment of the uncertainties of archives themselves, making the book more of a *heteroglossia* than a classical anthology with its more unified collection (p. 13). The term heteroglossia stems from the Russian literary theorist and linguist Mikhail Bakhtin[1] and describes the co-existence of and conflict between varying ways of evaluating, conceptualising, and experiencing the world. To Bakhtin, it is in the lingual reflection of "specific points of view on the world, forms for conceptualizing the world in words, specific world views, each characterized by its own objects, meanings and values"[2] that the radical potential of the heteroglossia is found. A potential to undermine *centripetal* (homogenising, hierarchising) forces and tendencies of language and culture through an exploitation of the inherent *centrifugal* (decrowning, decentering, dispersing) forces of language.[3] A similar centrifugal force is spinning through the *Uncertain Archives*. Instead of reinforcing their critique with a strongly linear and standardised structure, the editors have actively chosen to embrace the multiplicity of ways in which the uncertainty of big data archives can be identified and conceptualised in current scholarship and beyond.

As such, prescient methods for "hashtag archiving" (Tara L. Conley) are listed side by side with recovering of "absent presence" in large datasets (Lisa Blackman), and critical examination of the distancing effect of digital detox practices (Pepita Hesselberth)

side by side with contemplations on the entanglements of technological demos and the world they make visible and, ultimately, knowable, and debatable (Orit Halpern). The alphabetical ordering of the keyword terms allows for many voices, perspectives, and styles, and it invites the reader to approach the book in multiple ways. Yet, the format chosen for *Uncertain Archives* also demands a lot of its reader. To the curious student, for instance, an A-to-Z introduction to the emerging field of critical and cultural data studies could have made the volume easier to approach. Similarly, a more cohesive narrative about the ways in which large data archives shape our societal and political realities could have helped strengthening the much-needed critique running through the chapters.

Despite its challenging format, the book provides rich insights for the committed reader. As a nonhierarchical, dialogical, and radically *open* heteroglossia of big data knowledges, it practices what it preaches. Far from the viewpoint of one privileged observer, the coming-together of diverse voices, positions, and styles is, at its best, Harawayean in its ambition: embodied, feminist, and deeply situated in the world it portrays.[4] "It is our argument that big data must be analyzed from a range of different disciplinary vantage points—not least from the perspective of the humanities because big data interact at every level with the human," the editors write (p. 1). As a reaction to the Silicon Valley-driven big data *pharmakon*, this is a much-needed and highly recommended contribution.

Maja Bak Herrie

1   Bakhtin introduced the concept in his 1934 paper "Слово в романе [Slovo v romane]," published in English as "Discourse in the Novel." The English translators Michael Holquist and Caryl Emerson introduced the word *heteroglossia* as their translation of Bakhtin's term разноречие (raznorechie, meaning "varied-speechedness") in the book *The Dialogic Imagination: Four Essays by M. M. Bakhtin* (Austin: University of Texas Press, 1981).

2   Bakhtin, *The Dialogic Imagination*, 291-292.
3   Bakhtin, 272.
4   Donna Haraway, "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective," *Feminist Studies* 14, no. 3 (1988): 575-599. https://doi.org/10.2307/3178066