

Book Review: Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition by Wendy Hui Kyong Chun

In Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition, Wendy Hui Kyong Chun explores how technological developments around data are amplifying and automating discrimination and prejudice. Through conceptual innovation and historical details, this book offers engaging and revealing insights into how data exacerbates discrimination in powerful ways, writes David Beer.

***Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition.* Wendy Hui Kyong Chun (mathematical illustrations by Alex Barnett). MIT Press. 2021.**

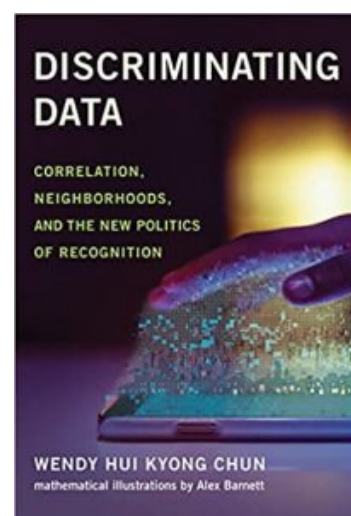
Find this book (affiliate link): [amazon](#)

Going back a couple of decades, there was a fair amount of discussion of ‘the digital divide’. Uneven access to networked computers meant that a line was drawn between those who were able to switch-on and those who were not. At the time there was a pressing concern about the disadvantages of a lack of access. With the massive escalation of connectivity since, the notion of a digital divide still has some relevance, but it has become a fairly blunt tool for understanding today’s extensively mediated social constellations. The divides now are not so much a product of access; they are instead a consequence of what happens to the data produced through that access.

With the escalation of data and the establishment of all sorts of analytic and algorithmic processes, the problem of uneven, unjust and harmful treatment is now the focal point for an animated and urgent debate. Wendy Hui Kyong Chun’s vibrant new book [Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition](#) makes a telling intervention. At its centre is the idea that these technological developments around data ‘are amplifying and automating – rather than acknowledging and repairing – the mistakes of a discriminatory past’ (2). Essentially this is the codification and automation of prejudice. Any ideas about the liberating aspects of technology are deflated. Rooted in a longer history of statistics and biometrics, existing ruptures are being torn open by the differential targeting that big data brings.

This is not just about bits of data. Chun suggests that ‘we need [...] to understand how machine learning and other algorithms have been embedded with human prejudice and discrimination, not simply at the level of data, but also at the levels of procedure, prediction, and logic’ (16). It is not, then, just about prejudice being in the data itself; it is also how segregation and discrimination are embedded in the way this data is used. Given the scale of these issues, Chun narrows things down further by focusing on four ‘foundational concepts’, with correlation, homophily, authenticity and recognition providing the focal points for interrogating the discriminations of data.

It is the concept of correlation that does much of the gluing work within the study. The centrality of correlation is a subtext in Chun’s own overview of the book, which suggests that ‘*Discriminating Data* reveals how correlation and eugenic understandings of nature seek to close off the future by operationalizing probabilities; how homophily naturalizes segregation; and how authenticity and recognition foster deviation in order to create agitated clusters of comforting rage’ (27). As well as developing these lines of argument, the use of the concept of correlation also allows Chun to think in deeply historical terms about the trajectory and politics of association and patterning.



For Chun the role of correlation is both complex and performative. It is argued, for instance, that correlations ‘do not simply predict certain actions; they also form them’. This is an established position in the field of critical data studies, with data prescribing and producing the outcomes they are used to anticipate. However, Chun manages to reanimate this position through an exploration of how correlation fits into a wider set of discriminatory data practices. The other performative issue here is the way that people are made-up and grouped through the use of data. Correlations, Chun writes, ‘that lump people into categories based on their being “like” one another amplify the effects of historical inequalities’ (58). Inequalities are reinforced as categories become more obdurate, with data lending them a sense of apparent stability and a veneer of objectivity. Hence the pointed claim that ‘correlation contains within it the seeds of manipulation, segregation and misrepresentation’ (59).

Given this use of data to categorise, it is easy to see why *Discriminating Data* makes a conceptual link between correlation and homophily – with homophily, as Chun puts it, being the ‘principle that similarity breeds connection’ and can therefore lead to swarming and clustering. The acts of grouping within these data structures mean, for Chun, that ‘homophily not only eases conflict; it also naturalizes discrimination’ (103). Using data correlations to group informs a type of homophily that not only misrepresents and segregates; it also makes these divides seem natural and therefore fixed.

Chun anticipates that there may be some remaining remnants of faith in the seeming democratic properties of these platforms, arguing that ‘homophily reveals and creates boundaries within theoretically flat and diffuse social networks; it distinguishes and discriminates between supposedly equal nodes; it is a tool for discovering bias and inequality and for perpetuating them in the name of “comfort,” predictability, and common sense’ (85). As individuals are moved into categories or groups assumed to be like them, based upon the correlations within their data, so discrimination can readily occur. One of the key observations made by Chun is that data homophily can feel comfortable, especially when encased in predictions, yet this can distract from the actual damages of the underpinning discriminations they contain. Instead, these data ‘proxies can serve to buttress – and justify – discrimination’ (121). For Chun there is a ‘proxy politics’ unfolding in which data not only exacerbates but can also be used to lend legitimacy to discriminatory acts.

As with correlation and homophily, Chun, in a particularly novel twist, also explores how authenticity is itself becoming automated within these data structures. In stark terms, it is argued that ‘authenticity has become so central to our times because it has become algorithmic’ (144). Chun is able to show how a wider cultural push towards notions of the authentic, embodied in things like reality TV, becomes a part of data systems. A broader cultural trend is translated into something renderable in data. Chun explains that the ‘term “algorithmic authenticity” reveals the ways in which users are validated and authenticated by network algorithms’ (144). A system of validation occurs in these spaces, where actions and practices are algorithmically judged and authenticated. Algorithmic authenticity ‘trains them to be transparent’ (241). It pushes a form of openness upon us in which an ‘operationalized authenticity’ develops, especially within social media.

This emphasis upon the authentic draws people into certain types of interaction with these systems. It shows, Chun compellingly puts it, ‘how users have become characters in a drama called “big data”’ (145). The notion of a drama is, of course, not to diminish what is happening but to try to get at its vibrant and role-based nature. It also adds a strong sense of how performance plays out in relation to the broader ideas of data judgment that the book is exploring.

These roles are not something that Chun wants us to accept, arguing instead that ‘if we think through our roles as performers and characters in the drama called “big data,” we do not have to accept the current terms of our deployment’ (170). Examining the artifice of the drama is a means of transformation and challenge. Exposing the drama is to expose the roles and scripts that are in place, enabling them to be questioned and possibly undone. This is not fatalistic or absent of agency; rather, Chun’s point is that ‘we are characters, rather than marionettes’ (248).

There are some powerful cross-currents working through the discussions of the book's four foundational concepts. The suggestion that big data brings a reversal of hegemony is a particularly telling argument. Chun explains that: 'Power can now operate through reverse hegemony: if hegemony once meant the creation of a majority by various minorities accepting a dominant worldview [...], now hegemonic majorities can emerge when angry minorities, clustered around a shared stigma, are strung together through their mutual opposition to so-called mainstream culture' (34). This line of argument is echoed in similar terms in the book's conclusion, clarifying further that 'this is hegemony in reverse: if hegemony once entailed creating a majority by various minorities accepting – and identifying with – a dominant worldview, majorities now emerge by consolidating angry minorities – each attached to a particular stigma – through their opposition to “mainstream” culture' (243). In this formulation it would seem that big data may not only be disciplinary but may also somehow gain power by upending any semblance of a dominant ideology. Data doesn't lead to shared ideas but to the splitting of the sharing of ideas into group-based networks. It does seem plausible that the practices of targeting and patterning through data are unlikely to facilitate hegemony. Yet, it is not just that data affords power beyond hegemony but that it actually seeks to reverse it.

The reader may be caught slightly off-guard by this position. Chun generally seems to picture power as emerging and solidifying through a genealogy of the technologies that have formed into contemporary data infrastructures. In this account power seems to be associated with established structures and operates through correlations, calls for authenticity and the means of recognition. These positions on power – with infrastructures on one side and reverse hegemony on the other – are not necessarily incompatible, yet the discussion of reverse hegemony perhaps stands a little outside of that other vision of power. I was left wondering if this reverse hegemony is a consequence of these more processional operations of power or, maybe, it is a kind of facilitator of them.

Chun's book looks to bring out the deep divisions that data-informed discrimination has already created and will continue to create. The conceptual innovation and the historical details, particularly on statistics and eugenics, lend the book a deep sense of context that feeds into a range of genuinely engaging and revealing insights and ideas. Through its careful examination of the way that data exacerbates discrimination in very powerful ways, this is perhaps the most telling book yet on the topic. The digital divide may no longer be a particularly useful term but, as Chun's book makes clear, the role data performs in animating discrimination means that the technological facilitation of divisions has never been more pertinent.

Note: This article first appeared at our sister site, [LSE Review of Books](#). It gives the views of the author, not the position of EUROPP – European Politics and Policy or the London School of Economics. Featured image credit: [Pixabay](#)
