Informational selves – a genealogy of how data became to increasingly define how we operate

Koopman, Colin. 2019. *How we became our data – A Genealogy of the informational person.* Chicago: University of Chicago Press. pp. 272 ISBN: 9780226626444.

Reviewed by Johannes Lenhard, Centre Coordinator and Research Associate, Max Planck Cambridge Centre for the Study of Ethics, the Economy and Social Change, University of Cambridge, UK

Just in February, a big blue Metropolitan police van was parked outside the exit of the London underground at the second biggest shopping center in the British capital in Stratford. Two big white cameras sat high on top of the van pointing at the outcoming crowds; a red sign with white letters was placed in proximity of the car but barely readable for people approaching. The sign read: 'Police Live Facial Recognition in Use.' Despite protests from groups such as Big Brother Watch, London police force announced shortly before this first sighting of the blue van that it would increasingly use facial recognition software in London to 'find and stop criminals.' How does that work? In fact, the process once the pictures are captured is simple: the identified faces are fed into a big database of acutely searched-for supposed criminals. In case of a match between a filmed face and the database, police captures the suspect immediately, so the theory.

Unfortunately, an independent review during trials of the software has found significant flaws with the practice of this data-driven prosecution aide. Researchers around Professor Peter Fussey

at the University of Essex have observed issues both with the lawfulness of the process and its effectiveness. The reviewers concluded that there is no explicit and only insufficient implicit legal authority for the use of facial recognition in domestic law on the one hand; they found on the other hand, that the machine-identified matches during the trial period were only in 8 out of 42 cases 'verifiably correct.' That is a terrible 20% quota.

***

The use of facial recognition software by the police in London (and elsewhere – see conversations in <u>California</u>) opens up questions of both ethics and data reliability that – as Koopman shows in his excellent genealogy of the 'informational person' – date back at least 100 years. Who can legitimately collect which kinds of data? What kind of usages for the data once collected are legitimate? What kind of models can help make use of the data and which assumptions are fed into these models? Do the models lead to reliable results with regards to the set goals? Koopman starts even more simple than that, however, and asks: how, in what formats is data collected?

He traces back many of these highly topical questions to the 1910s and, in a very explicitly Foucauldian genealogical analysis, he begins with a startling observation: "our information composes significant parts of our very selves." (vii). His fundamental curiosity: "how have [we] become our data?" (ix). He argues, to put this differently, that we are 'informational persons' "inscribed, processed, and reproduced as subjects of data" (4) (to a point where we would lose a significant part of our selves if we lost our data, 5) and wonders: when and how did that start?

With a strong focus on the American context, Koopman presents three very detailed case studies of the traces he found about 120 years ago: the beginning of identity documents (birth certificates and national insurance number), the origin of personality traits (in personality testing) and the history of racialized real-estate redlining. In the first case study, Koopman puts forward a very vivid example of a 'form' and a 'format' that fixes everyone from the moment they are born and hence marks the "first point of entry into the information systems" (41). The practice of issuing birth certificates started off in the US as late as 1905 with a first bill in Pennsylvania

which was subsequently copied into other states. But despite a strong audit culture around the certificate (driven by the Census Bureau and so-called 'Birth Registration Areas,' 54), it only found true universal application when, with the social security number rollout in the mid-1930s, a strong incentive was established (61). The birth certificate became the necessary proof of age to be able to work and to receive social security benefits and later pension payments. On the flip side, the state was able to measure tax incomes and track wages. As Koopman claims: "SSN solidified […] a stable kind of person who could move to the beat of the drum of data" (63).

In the second case study, Koopman goes into detail on how personality has become a vector of ourselves as something "measurable and differentiable" through measuring personality traits (69) from the 1930s onwards. While developed scientifically as part of the field of 'abnormal psychology,' it was again the state that made intelligence and psychometric tests a standard measure of personality traits (93) and eventually character, starting with their WWI recruits (77). Koopman goes into much detail about the different kind of test regimes and forms as well measurement of personality more generally, but – just as with the first case study – one thing I am missing is not just a genealogy of the tests as such but their politics. What are these tests *not* measuring? Do they claim 'objectivity' and if so what is this based on? More importantly what consequences do the underlying assumptions have on the data collected, and eventually on the processed outputs?

The third and by far most detailed case study on racialized real-estate appraisal methods – what is often described as redlining – is in contrast focused strongly on this aspect of the informational person: the ethics and politics of data gathering, processing and output. Koopman lays out how real estate appraisals – judging the price of a given property – became scientific-ized in the 1920s mostly driven by increasingly professionalizing real estate brokers (122). The problem was that one of the inputs in the appraisal method was "new forms of racialized evidence" (116); in simple words: in order to fulfil their 'economic duty' and value the property correctly, the 'racial composition of the neighborhood' had to be considered (119). Race was a "valuation factor" (126) that was worked into standard valuation work sheets and reports (128f) based on guidelines published in newly conceptualized 'manuals' (133). This had immediate effects not only on house prices but also on mortgages (and the possibility of remortgaging properties) as

the Home Owners' Loan Corporation (HOLC), which gave out specific kinds of mortgages, was among the first to 'redline' maps in 1933 (138). While their goal was to extend access to mortgages, they were trying to de-risk their operations by grading properties and areas on a scale of 'residential security': "black neighborhoods were invariably rated as Fourth grade," i.e. the worst possible grade (141). Naturally, these techniques produced highly unequal decisions negatively affecting access to properties and credit for African-Americans and – as Koopman does point out – often continue to do so.

***

What Koopman calls technical formatting is an act of power that fastens its subject to their data (12), all in the name of efficiency and better measurement and decision making; ultimately, this kind of analysis based on standardized forms is supposed to be more scientific, quicker, more reliable and more objective. But the consequences of these processes and the related infopower as Koopman terms it can "limit our freedoms or open up our liberties" (159). While Koopman seems to stay mostly neutral as to which direction – limiting or opening up liberties? – the processes enabled more, as I already suggested above, he closes his volume with a question implicitly fearing the danger of the techniques: "what can we do to redesign these formats where they function today as programs for the disposition of burdens?" (180).

Koopman is doing something very well that many other authors – partly because of their disciplinary affiliation, partly because it is a really tedious task – don't do: he goes back in history to trace the origins of something we are living with today – the fixation as our beings as informational. What remains for me beyond my quarrel with his selective digging deeper into the politics of the techniques is twofold: first, the state throughout Koopman's analysis plays a crucial role, as financier, as regulator, as administrator of the data collection, procession and validation; perhaps a further step in the analysis could zoom in on the role the state played in the development of the informational self more explicitly (inspired by authors such as Mazzucato 2017, for instance). Secondly, I am curious to hear more directly about how he thinks the issues he unearths with data – from its collection, its procession to the outputs it generates – relate to problems today. What exactly can we learn from the teething issues of the birth certificate? How

are the historical formats still shaping outputs today? There are surely other authors (e.g. O'Neil 2016; Noble 2018) who pick up this thread very elegantly. Obviously, as a European, I also now want to see Koopman's American genealogy replicated to this side of the planet – and I hope somebody as careful and considerate as Koopman takes that task up in the near future.

Dr. Johannes Lenhard is the Centre Coordinator of the Max Planck Cambridge Centre for the Study of Ethics, the Economy and Social Change (Max Cam). While his doctoral research was focused on the home making practices of people sleeping rough on the streets of Paris, he is currently working on a new project on the ethics of venture capital investors in Berlin, London and San Francisco. Find his twitter at @JFLenhard.