16.   The Myth of Big Data

Nick Couldry

Increasingly, institutions in the fields of research and policymaking, as well as the corporate realm, base decision-making and knowledge production on metrics calculated from what is metaphorically called 'big data'. In his 2013 inaugural lecture 'A Necessary Disenchantment: Myth, Agency and Injustice in a Digital World', Nick Couldry discusses the mythical claim that big data is generating a new and better form of social knowledge.

What does/doesn't 'big data' tell us about the social? Or in other words, what type of 'social' is being constructed in social media?

No one disputes that data sets in every domain, including those relating to the social, are very large, or that, because they are so large, there is something to be gained by using automated processing to establish correlations across those data sets; such processing is, of course, beyond the capacity of human interpreters. The issue is how we interpret the value of the outputs of such processing. Already in the latently metaphorical term big data, there is a story being told about human beings' changing relation to the domain we have called social that is highly contestable. Big data, it is implied, is the source of a different order of knowledge, a step change in human self-understanding that precisely bypasses humans' meagre attempts at self-understanding through interpreting the local details of what they think, say and do. This way of putting things obviously prejudges positively the value of the outputs of 'big data' processing. That is the ideological work done by the term 'big', beyond its obvious descriptive force (and, as I said, no one disputes that the data sets involved are very large!). I don't believe we should accept this story, and I will come back later to how we could contest it. But if we simply accept it, it has major consequences for the type of social domain that is accessible to us as researchers and social actors. 'Big data' is only possible on two basic conditions (which actually are composites of many more detailed conditions): first, that data is collected continuously about the states of affairs in various domains (including not just what individuals do and say, but the state of their bodies); second, that data is aggregated and its patterns of correlation computed and 'interpreted'. Because only information of particular sorts conforms to the requirements of data management, and because only processes of particular

sorts generates such information, 'big data', however expanded its scope and however fine-tuned its workings, must always be a selection from the actual world of action and interaction. danah boyd and Kate Crawford (in their important 2013 article on the 'myths' of big data) brought out the many specific delusions in relying on 'big data' as a source of knowledge, but in my work I have tried to focus on the overall delusion that, if you like, frames all the specific ones: that is, the overall attempt to reorient us towards big data processes as 'the' new form of social knowledge.

Ideological uses of the term 'big data' however forget that general, yet highly motivated, selectivity, and so inevitably misread the picture of the social obtained through big data processing, but with a constructive force that is difficult to resist, especially when investment in social knowledge (by governments, funders, private corporations) is increasingly focused on 'big data'. Over time, this may obscure our possibilities for imagining, describing and enacting the social otherwise. Meanwhile, the ideological work is going on all around us, whether in Wired magazine editor Chris Anderson's trailblazing article 'The End of Theory' (Anderson 2007) or more critically in a book such as Mayer-Schönberger and Kenneth Cukier's *Big Data: A Revolution that Will Transform How We Live, Work, and Think* (2013) where they predict that, as datafication and big data processing grows, 'we will no longer regard our world as a string of happenings that we explain as a natural or social phenomenon, but as a universe comprised essentially of information' (2013: 96, emphasis added). Note that opposition: not 'natural or social' but 'essentially information'.

Clearly those with skills at interpreting the social 'need not apply' in this new world of 'social' knowledge! The paradox, as with most forms of symbolic violence, lies only just below the surface, but it relies for its effectiveness on us letting it pass without comment – and on us acting out its consequences every day. I will come back to action and resistance in a moment. The long-term consequences of this ideological shift towards 'big data' as the new default source of knowledge about 'the social' takes two contrasting forms.

First, in how the particular details of data collection and data processing recalibrate the possibilities of social existence, the ontology if you like of the social. We feel this at work from hour to hour as we monitor our day to decide whether it is 'worth' a status update on whatever social media platform we use. But the detailed workings are much harder to track, and require exhaustive analysis of the linked data sets on which, for example, automated credit ratings are based. Oscar Gandy was pioneering in seeing the socially discriminatory potential of corporate data collection a quarter

of a century ago (Gandy 1993), and my LSE colleague Seeta Gangadharan is doing great work in this area: www.datacivilrights.org/.

And second – and this was more my focus in my inaugural lecture, at LSE (Couldry 2014) – to criticize the overall celebration of social knowledge achieved through automated, processing and, by contrast, the devaluing – even the attempted decommissioning,– of other forms of social knowledge that until fairly recently, were taken seriously. To unpack this, we need 'a hermeneutic of the anti-hermeneutic'; we need to register what Judith Butler calls 'a refusal of discourse' (2004: 36). If we don't, we risk losing touch – in our languages about the social – with a basic truth: that, as philosopher Charles Taylor (1986) put it, the human being is 'a self-interpreting animal' and so the only possible meaning of our lives together stems from its basis in our attempts to interpret what we do to each other. Without a hold on that truth, we accept a risk of inhabiting what the 19th-century Russian novelist Nikolai Gogol called 'dead souls': human entities that have financial value (in his novel, as mortgageable assets; in our new world, as unwitting data producers), but that are not alive, not at least in the sense we have always known human beings to be alive.

But why talk about the claims about big data as mythical? Why should we care so much about the myth of big data?

That goes back to the question of action. In much of my work I have been concerned with how it is that large modern societies have become organized, if you like focused, around the productions of particular institutions with huge symbolic power, whether traditional media institutions or increasingly the organizations that run our digital platforms and also those that generate process and own the data that we, largely unwittingly, generate through our actions online. I have always argued that such a big social 'fix' requires something more than ideology in the traditional sense: it requires us to act in ways that conform to it. Here Žižek's concept of ideology (1990) is more helpful than Marx's, but personally I have preferred to use the term 'myth'. The myth of big data is a fix of that sort: a society-wide rationalization of a certain state of affairs that works not just, or sometimes not even, through what we think, but always through what we do: what we go on doing, whatever we believe ('clicking like' and so on). The myth of big data is particularly broad in how it has emerged and is being played out, but is also particularly important in that it works to challenge the very idea that the social is something we can interpret at all. It works to disable other, older (and no doubt newer) forms of social knowledge. That, I believe, will have

huge consequences in the longer term for our understandings of democratic agency and social justice.

So how then do we achieve a more agent-focused account of big data?

This is a long-term collective battle, and there are many levels to it. First, we need to refuse the myth as such: to reject its explicit claims and language. The myth of big data is an attempt to appropriate the possibilities of producing knowledge about the social domain, which needs to be resisted. So specific attempts to claim better understanding of the social based on automated processing of very large data sets need to be closely interrogated as specific claims, stripping away the usual rhetoric about 'how all social knowledge is changing' that often accompanies such specific claims.

Second, and as another LSE colleague, Alison Powell, and I argued in an article called 'Big Data from the Bottom Up', the skills and collaborative practices necessary for those outside the large institutions that benefit from the myth of big data – including civil society organizations – to work with large data sets must be developed and encouraged. It is the case – and this is the good element in some big data rhetoric – that cities might be run better if citizens gathered different types of data about what goes on in cities, and were empowered not only to decide how that data is being analysed to citizens' mutual advantage, but also what sorts of action might flow from the knowledge the analysis of that data generates. Part of that process of opening up civically the black box of 'big data' (but which we would do better to simply call very large data sets) means spreading awareness of how currently vast data sets are collected, sifted, aggregated and then repackaged as sources of truth, but without much, if any, accountability for the rules of operation that drive that process.

This practical civic project holds to a basic principle of social science research, that it should work towards the 'de-reification' of social processes (Sewell 2005). Never have we needed a project of 'de-reification' more than today, I suspect.

References

Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *WIRED*. 23 June. www.wired.com/2008/06/pb-theory/.

boyd, danah, & Crawford, Kate. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, And Scholarly Phenomenon." *Information, Communication & Society* 15 (5): 662-679.

Butler, Judith. 2006. *Precarious Life: The Powers of Mourning and Violence*. Verso.

Gandy Jr., Oscar H. 1993. *The Panoptic Sort: A Political Economy of Personal Information. Critical Studies in Communication and in the Cultural Industries.* ERIC.

Gogol, Nikolay. 2004. *Dead Souls*. New York: Penguin Press.

Mayer-Schönberger, Viktor & Kenneth Cukier. 2013. *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.

Sewell Jr, William H. 2005. *Logics of History: Social Theory and Social Transformation*. Chicago: University of Chicago Press.

Taylor, Charles. 1985. "Self-Interpreting Animals." In Taylor, Charles: *Human Agency and Language*. Cambridge: Cambridge University Press, 45-76.

Žižek, Slavoj. 1989. *The Sublime Object of Ideology*. London: Verso.