The taming of chaos: Optimal cities and the state of the art in urban systems research

Linnet Taylor
Tilburg University, the Netherlands

Abstract
What can urban big data research tell us about cities? While studying cities as complex systems offers a new perspective on urban dynamics, we should dig deeper into the epistemological claims made by these studies and ask what it means to distance the urban researcher from the city. Big data research has the tendency to flatten our perspective: it shows us technology users and their interactions with digital systems but does so often at the expense of the informal and irregular aspects of city life. It also presents us with the city as optimisable system, offering up the chance to engineer it for particular forms of efficiency or productivity. Both optimisation itself, and the process of ordering of the city for optimisation, confer political and economic power and produce a hierarchy of interests. This commentary advocates that researchers connect systems research to questions of structure and power. To do this requires a critical approach to what is missing, what is implied by the choices about which data to collect and how to make them available, and an understanding of the ontologies that shape both the data sets and the urban spaces they describe.

Keywords
big data, urban systems research, mobile phones, transport, optimisation

Received March 2021; accepted March 2021
Commentary

The earliest photographic images of cities are notable for their lack of people. The urban landscape plays a starring role in the work of photography pioneers such as Daguerre, Fox Talbot and Bayard, with city scenes that are at once strikingly recognisable and alien in their stillness. They are rich with meaning, indicating conditions of poverty or growth and prosperity, the reconstitution of public space by development, and sometimes war and conflict. Yet they are portraits of absence. When Bayard photographs the Paris commune of 1848, no people are visible (see Figure 1). Even in Thibault’s image of the barricades immediately following an attack, people are only present as a ghostly blur (see Figure 2). In all these images, the technical achievement of recording the urban landscape substitutes for explaining its human dynamics. But this is not a choice: people moved too fast to be portrayed and had to be photographed in individualised sittings where they were often held still by clamps so that they could be captured by the exposed film. People were incidentally present in urban landscapes if they stayed still long enough, but the early photographer had to choose between representing the city and representing its inhabitants.

The state of the art in big data urban research shares some features with this early photography. Just as Daguerre and his peers were working in a technical domain that was effectively less than two decades old, we are entering the second decade of widespread urban systems research. And, as with early photography, the new methods and epistemologies offered by big data analytics provoke public wonder at their experimentation and exploration. This makes the current moment one where we can surface and render debatable these still-developing urban

Corresponding author:
Linnet Taylor, Tilburg Institute for Law, Technology and Society, Tilburg University, PO Box 90153, Tilburg, 5000 LE, the Netherlands.
Email: L.E.M.Taylor@tilburguniversity.edu
research epistemologies. This field of research traces how apparently familiar dynamics operate at different and larger scales, and makes the claim that it is both possible and relevant to study cities as complex systems. It is important not to accept these claims at face value but to interrogate them while they are still new. It is not a given that tracing the paths and practices of urban life as ‘the Eagle and the Wild Goose See It’, in the words of the first US aerial photographer JW Black, or with ‘the God’s eye view’ (Pentland 2012: 11), brings new insights, rather than changing the scale and methods of urban research. It is worth asking what this new distancing of the researcher from the object of study does, and what it means to acquire a god’s-eye view of the city.

Kitchin’s (2014) claim that large-scale, born-digital data sets bring with them their own ontologies remains salient nearly a decade after it was made. Computational ontologies shape which data are collected and from whom, and therefore what researchers can see of the world by using those data. Shearmur (2015: 3) warns, more specifically, that ‘Big Data are not about society, but about users and markets’. Urban research conducted using born-digital large-scale data therefore could be said to sample from a subsample, the representativeness of which is uncertain. Urban systems research can represent dynamics in the city, often with unprecedented granularity and scale, but whose city remains unclear – the people who form the systems in question are present but invisible, seen only as an absence. Like traces of light on silver nitrate, their presence is implied but not explained. This is why Kitchin refers to data-driven (urban) science as discursive: it is open to different ascriptions of meaning depending on one’s standpoint.

Urban systems analysis is helpful in that it lays bare not only the computational but also the philosophical ontologies that underpin big data research. ‘Culture, politics, policy, governance and capital’ (Kitchin, 2014: 5), are missing from the big data view. Instead, these are replaced by ‘paths, nodes, districts, edges and landmarks’ (Wang and Vermeulen, 2021: 3120), the semiotic features of datafied urban space. There are clearly many things that such an analysis cannot capture. Rather than assume systems research is lacking, however, a more interesting analysis might focus on this as a relational problem. What can, and cannot, urban systems research tell us about the world, and do these things counterbalance each other?

First, it tells us that it is becoming possible to study cities all over the world using the same big data analytical tools. It is no longer an issue that by doing so we focus exclusively on high-income countries or what Dalla Corte et al. (2017) have termed ‘instrumented’ cities – cities that were either designed to consume and produce digital data in the course of their operations, or that have been made datafied by the addition of sensors and analytics systems. However, using big data to analyse cities in low- and middle-income countries often has the effect of flattening the notion of what constitutes city life. The more extensive people’s engagement with the city – and the city’s with its people – through technology is, the more multidimensional the data that can be collected by observing the by-products of digital systems. If most people in a city lack smartphones, or the city’s digital infrastructure is oriented toward the needs of a minority, correspondingly the data the city emits may be complex and extensive but it will be thin in terms of its value for understanding the city. For example, when Bangalore datafied its water distribution system with sensors and controls on irregular use, a particular narrative reflecting sustainability and formal access became available through the digital by-products of the commercial system it
installed. Conversely, a completely different alternate narrative was available from conversations with the residents of informal neighbourhoods, who experienced a correspondingly less serviceable and organised water system, and a different definition of sustainability, than the one conveyed by a ‘big data’ account of the system (Taylor and Richter, 2017). Each perspective, computational or ethnographic, has its own ground truth, and these ground truths are mutually inconsistent.

This suggests that a comparison of systems research with ground-level observation might tell us something useful about imbalances in political representation, infrastructural equity and the sustainability of the systems supporting urban life. These two visions become particularly unbalanced in cities with high inequality: a study of segregation using transport data in the USA (Candipan et al., 2021) is possible and relevant because it forms part of a broad landscape of research that demonstrates and analyses segregation from the human as well as the systems perspective. In Myanmar or Xingjiang, in contrast, segregation is also enforced through infrastructure and visible in data, but those sources of data are politically closed and this view is inaccessible. The only ground truth there comes from interviews with those on the receiving end of injustice, making it impossible to build on this individual-level experiential reality with system-level studies.

Comparing the findings these different methodological worldviews can produce is useful because missing data are also data. The poor and marginalised do not leave fewer digital traces but qualitatively different and less accessible ones. If the lives of the free and wealthy can be observed through data on gentrification, transport and their consumption of goods, then the lives of the poor and constrained can be mapped through digital policing systems and governmental anti-fraud analytics, through digital traces of residential and educational segregation, through mapping air quality and outdated infrastructure, through cash transfers and commercial redlining systems, and through the use of short-term financial products such as airtime purchases and payday loans. Everyday behavioural data on work and mobility amongst the poor can be found in delivery-tracing apps and on gig economy platforms, in hospital admissions and court judgements, and in data on forced migration. They are never missing from the picture: instead, researchers’ methodological choices inevitably make one group or the other visible as a presence or as an absence.

Does this matter? Could we not just say that researchers are constrained by the availability of data? Does research always need to acknowledge what it is not showing and what roads were not taken? And does this question not apply just as much to studies of poverty and injustice?

If the technology were neutral, none of this would matter. The choice of data set, city or problem would not have such implications in terms of how cities develop and in whose interests. Yet the datafication of urban systems and infrastructure, as well as the growing opportunities for digital urban research, clearly have implications for what Shaw and Graham have termed ‘our digital rights to the city’ (Shaw and Graham, 2017).

One chief functional characteristic of the datafied city is its availability (and, by connection, that of its residents) to optimisation. For example, if the flow of people through an urban transport system (e.g. Liu and Miller, 2021) can be mapped with streaming data, those flows can also be reshaped to achieve particular goals by adjusting infrastructure, pricing or accessibility. These could include making the average commute from the suburbs faster or minimising the number of connections that people coming from poorer neighbourhoods have to make.
to get into the centre. As this indicates, optimisation is profoundly political because when a system is optimised, a choice is made to emphasise one type of outcome at the expense of others, which means selecting for the interests of one group over another.

Another example is offered by Kulynych et al. (2020) in their analysis of Waze, a digital service which directs drivers to roads with less traffic. These roads, the authors point out, tend to be in residential neighbourhoods. This kind of optimisation in the interest of rush-hour commuters comes with consequent effects on those neighbourhoods, including pollution and decreased safety on formerly quiet streets. In a city where Waze’s system-level vision of the optimal traffic distribution prevails, it eventually becomes necessary to redesign residential neighbourhoods with new infrastructure to calm traffic, control pollution and dissuade speeding. Waze also comes with the underlying assumption that fundamental policy change is not an option, and that cars must remain the primary mode of commuter transport. The discourse of optimisation, unlike that of a genuinely open policy process, tends to favour reshaping the status quo rather than challenging the premise of the question.

This is important to bear in mind because research based on big data tends to acquire a policy-related diagnostic function, regardless of the authors’ intentions. It comes ready-made as policy advice because it singles out what can be quantified and observed at scale, and therefore what is accessible for shaping by technical interventions. ‘Paths, nodes, districts, edges and landmarks’ can be addressed cleanly using a technical problem-framing in ways that messy social life cannot. Moreover, research that makes visible possible targets of optimisation also plays a political role in determining what gets optimised. If research can ‘demonstrate how a smart city can be framed and implemented as a tool for tackling endogenous social challenges’ (Trencher, 2019), then it matters greatly what researchers choose to address and thus to make visible. Every big data analysis of urban systems inevitably frames the city as a laboratory and indicates possible experiments on urban life.

Given that datasets come with their own politics, this makes researchers’ choice of object and of methodological approach important in determining how urban governance and development are designed and executed. This also raises the question of whether research should address urban system dynamics as political from the start, that is, as a discourse that potentially opens up or closes off avenues of change and rethinking, or whether the researcher can stand back and allow others to add in the politics later as a kind of social scientific special sauce. At the least, it implies that researchers should be aware of the possible implications of their choices. A study that analyses the spatial and behavioural dynamics of a political protest, for instance, inevitably has the potential to shape law enforcement interventions. Analysing the dynamics of informal labour could allow authorities to protect irregular workers but could also suggest ways to suppress the informal economy on which many survive. Showing how people move through the city centre on a busy day helps authorities with designing public safety interventions. It also offers the potential for authorities to optimise pathways through the area to maximise retail consumership, or to render political protests harder to organise.

This is as much as to say that there is no neutral science when it comes to making visible urban systems. Cities are spaces of continual experimentation and redesign, and the analytics that inform those processes matter greatly for the kind of governance that is possible. Researchers have a degree of power to direct how their work will be
used: they can point to missing data, to the underlying logics of the systems they analyse, or to assumptions that might be worth questioning. For example, research on crime has the option of aligning with underlying carceral logics (Jefferson, 2017) by accepting the premise of predictive policing but also has the option of questioning that premise by mapping white collar crime, usually invisible, instead (Tseng et al., 2017). Similarly, researchers conducting big data studies on migration or integration must choose whether they accept the underlying framing of migration as security and economic risk that often comes with access to big data (Taylor and Meissner, 2020). Researchers are free to choose whether they frame existing infrastructure and resource inequities as inevitable or as stemming from political choices and, similarly, whether they use informality and irregularity as proxies for crime and risk or as signs of dysfunctional systems that are open to change.

These choices exist because increased granularity and volume of data do not ensure a thick description of urban systems. Instead, they require a critical approach to what is missing, what is implied by the choices about which data to collect and how to make them available, and an understanding of the ontologies that shape both the data sets and the urban spaces they describe. These questions are relevant regardless of the location of the city being studied, whether the framing is that of the Sustainable Development Goals or economic optimisation for the retail sector. They also imply something about the ‘who’ of urban systems research: it is difficult to understand the range of possible choices unless one is familiar with the city in question. Yet, those who can access and analyse systems-level data are rarely those familiar with the lived experience of a city – a digital divide already predicted in the early phase of big data research (boyd and Crawford, 2012; Savage and Burrows, 2007). These authors warned that big data are not useful without the permission, infrastructure, education and resources to manipulate them. These different gatekeeping points make critical research using big data both less likely and more important to consider on the part of those who do get access.

Control over data, and thus the questions that can be asked about cities, is a question of power and politics. It reflects both the dimensions of power over and power to (Lukes, 2004), helping to determine whether, and how, people can influence the governance and design of the city, and also their own agency. It is always worth asking why certain places, systems and dynamics can be made visible and others not – what cannot be optimised, or even addressed, often remains in the background while research focuses on accessible problems. The field of urban research therefore faces a new challenge of connecting a datafied view of the city to the lived reality of urban life and governance. Data at scale do not have to imply a remoteness from the politics of the city; instead, the challenge becomes to relate systems research to questions of structure and power.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This commentary was written with support from the Horizon 2020 program, ERC StG 757247.
Note
1. Black JW (1860) *Boston, as the Eagle and the Wild Goose See It*. Available at: https://www.metmuseum.org/art/collection/search/283189

References

boyd D and Crawford K (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.

Candipan J, Phillips NE, Sampson RJ, et al. (2021) From residence to movement: The nature of racial segregation in everyday urban mobility. *Urban Studies* 58(15): 3095–3117.

Dalla Corte L, van Loenen B and Cuijpers C (2017) Personal data protection as a nonfunctional requirement in the Smart City’s development. In: *Proceedings of the 13th International Conference on Internet, Law & Politics*. Barcelona, Universitat Oberta de Catalunya, 29–30 June. Available at: https://research.tilburguniversity.edu/files/23654477/Proceedings_IDP_17_1_.pdf

Jefferson BJ (2017) Digitize and punish: Computerized crime mapping and racialized carceral power in Chicago. *Environment and Planning D: Society and Space* 35(5): 775–796.

Kitchin R (2014) Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1): 205395171452848.

Kulynych B, Overdorf R, Troncoso C, et al. (2020) POTs: Protective optimization technologies. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* ’20)*, Atlanta, US, 27–30 January 2020, pp. 177–188. Available at: https://doi.org/10.1145/3351095.3372853

Liu L and Miller HJ (2021) Measuring risk of missing transfers in public transit systems using high-resolution schedule and real-time bus location data. *Urban Studies* 58(15): 3140–3156.

Lukes S (2004) *Power: A Radical View*. London: Macmillan International Higher Education.

Pentland A (2012) Society’s nervous system: Building effective government, energy, and public health systems. *IEEE Computer* 45(1): 31–38.

Savage M and Burrows R (2007) The coming crisis of empirical sociology. *Sociology* 41(5): 885–899.

Shearmur R (2015) Dazzled by data: Big Data, the census and urban geography. *Urban Geography* 36(7): 965–968.

Taylor L and Meissner F (2020) A crisis of opportunity: Market-making, big data, and the consolidation of migration as risk. *Antipode* 52(1): 270–290.

Taylor L and Richter C (2017) The power of smart solutions: Knowledge, citizenship, and the datafication of Bangalore’s water supply. *Television & New Media* 18(8): 721–733.

Tseng F, Clifton B and Lavigne S (2017) White collar crime risk zones. *The New Inquiry* 59 (March) Available at: https://whitecollar.thenewinquiry.com

Wang M and Vermeulen F (2021) Life between buildings from a street view image: What do big data analytics reveal about neighbourhood organisational vitality? *Urban Studies* 58(15): 3118–3139.