# Predictive Geographies*

BY MARK SHEPARD

Information today is embedded within, and distributed across, a variety of urban and exurban geographies. Indeed, it has become difficult to point to situations where information systems do not play a direct or indirect role in the disposition of physical features of the city and its surroundings, or of human activity as it affects and is affected by them. On any given day, we gain access to transportation systems using a magnetic strip card or radio frequency ID (RFID) tag to pay a fare. We coordinate meeting times and places through SMS text messaging on a mobile phone while on the run. We check in at our favourite venues using social media apps on our mobile phones, and leave tips for strangers about what to do there. We cluster in cafes and parks where Wi-Fi is free to check email on a laptop. We drive cars with on-board navigation systems that map optimal routes to a destination, updating in real-time to take into account current traffic congestion, construction activity, and special events. We pass through public spaces blanketed by CCTV surveillance cameras monitored by machine vision systems running advanced face detection and object tracking algorithms.

These geographies of information are increasingly driven by data. Data are the raw material from which they are formed, by which they are constituted, and through which they take shape.[1] These data take a variety of forms: samples, indexes, measurements, abstractions and representations of the physical world and the events and activities that transpire within it. They vary in size (big data, small data); are qualitative and quantitative in nature; are structured, semi-

1. For an historical perspective on how data came to be understood as being pre-analytical and pre-factual, existing prior to interpretation and argument, see Daniel Rosenberg, "Data Before the Fact." in *Raw Data Is an Oxymoron*, ed. Lisa Gitelman (Cambridge: MIT Press, 2013). For an overview of how this understanding has evolved over recent years in considering how data are constitutive of the ideas, techniques, technologies, people, systems and contexts that conceive, produce, process, manage, and analyze them, see Rob Kitchin and Tracey P. Lauriault, "Towards critical data studies: Charting and unpacking data assemblages and their work" (January 8, 2014). *The Programmable City Working Paper* 2, accessed March 5, 2015, http://ssrn.com/abstract=2474112

structured, and unstructured in organisation; and may be captured, produced, or derived through a variety of processes and procedures.[2] Our smart phones capture data on where we've been and produce location histories from which our mobility patterns can be derived. Our fitness activity trackers monitor a range of health factors – heart rate, steps taken, floors climbed, calories burned, even sleep quality – and produce representations of our progress toward self-identified goals that are shared and aggregated through online portals.

2. Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (London: Sage, 2014).

Data that information systems collect, transmit and store promise greater control over the performance of urban systems and new insights into how the city is inhabited collectively. Embedded systems monitor, manage, and regulate utility services and critical urban infrastructure. Camera networks monitor street intersections and issue traffic violations using automated license plate recognition software. Social media platforms like Twitter and Foursquare provide new ways of tracking how we move, whom we are with, where we go, and what we think of where we've been. Data on trending venues, popular neighbourhoods, peak times, and common frequencies enable us to map patterns of mobility and activity in a more dynamic and fine-grained manner. With the advent of big data, some claim we have the opportunity to connect, aggregate, analyse, and integrate data about the urban environment in ways that enable us to better visualise, model, and predict urban processes; simulate probable outcomes; and lead to more efficient and sustainable cities.[3]

3. Rob Kitchin, "The real-time city? Big data and smart urbanism." *GeoJournal 79* (2014): 1 – 14.

Yet data are relatively meaningless in raw form. From data, to information, to knowledge, to understanding, to wisdom: each step up the stack requires some form of processing (correlation, interpretation, abstraction, analysis, reduction, aggregation, etc.) that adds value to the previous step.[4] Small data, such as that produced through surveys, interviews or methods of direct observation common to the social sciences, has to date been processed primarily by people. Often designed to address specific research questions, small data focuses on specific cases and tells individual, nuanced and contextual stories.[5] With the advent of big data, algorithms are playing an increasing role in the process by which data becomes information. Data mining algorithms, for instance, operate on large data sets and look for patterns in data that are not readily apparent to people. As algorithms replace people in the derivation of in formation from data, the production of knowledge

4. Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences.*

5. Rob Kitchin and Tracey P. Lauriault, "Small Data, Data Infrastructures and Big Data" (January 8, 2014), in *The Programmable City Working Paper 1,* accessed March 5, 2015, http://ssrn.com/abstract=2376148 or http://dx.doi.org/10.2139/ssrn.2376148

becomes a process shared by human and non-human actors.

Algorithms also do work in the world. They play a key role in the constitution of particular physical spaces, such as airport check-in lounges or supermarket checkout aisles, for instance. These kinds of spaces rely on algorithms to perform their basic programmatic functions. When the information systems upon which these spaces depend fail, the spatial program transforms: an airport lounge becomes a waiting room, a supermarket aisle becomes a warehouse.[6] Algorithms also influence the organisation of space at urban and regional scales. Consider, for example, the role of high speed trading algorithms in the reorganisation of the financial sector of lower Manhattan. When algorithms running on computers located in high-speed trading hubs compete with each other, the latencies of electronic transactions measured in microseconds become critical. The distance between the servers running these algorithms and the trading hubs can make a difference measured in millions of dollars per second. As face-to-face trading is replaced by electronic trading, the value of location becomes more a function of proximity to high-speed communications backbone infrastructure than to centres of human interaction and exchange.[7]

6. Rob Kitchin and Martin Dodge, *Code /Space: Software and Everyday Life* (Cambridge: MIT Press, 2011).

7. Kazys Varnellis, "The Architecture of Financialization," *Perspecta 47* (2014) 185 – 194.

Matters become more complicated when algorithms gain predictive or anticipatory agency. As we move beyond the initial practices and promises of both the Smart City and the Smart Citizen, we find information processing capa-city approaching a higher order of magnitude.[8] These emerging information systems operate on aggregate datasets using predictive analytics. Predictive analytics refers to a variety of statistical techniques – modelling, machine learning, and data mining – that analyse current and historical facts to make predictions about future events. While predictive analytics is closely entwined with data mining, the inferences derived through predictive analytics extend beyond retrospective pattern analysis to incorporate more prospective and anticipatory postulations. Where data mining describes the exploratory process of finding patterns and information within data, predictive analytics attempts to leverage that information derived from data to anticipate meaning and make predictions about the future.

8. Mark Shepard, "Beyond the Smart City: Everyday Entanglements of Technology and Urban Life," *Harvard Design Magazine* 37 (2014): 18 – 23.

Predictive analytics has found widespread application in marketing. The *New York Times* published an article a few years ago that outlined the process by which big-box retailer Target deployed predictive analytics to identify women likely to be in their early stages of pregnancy so that it could target advertising to them before their child was born.[9] The algorithm that was behind the analytics was dubbed the "pregnancy-prediction algorithm". Predictive analytics is also at work in social media. The popular online dating website OkCupid uses a relatively simple algorithm based on self reported data – how you respond to a question, how you want your prospective partner to respond to the same question, and how important the question is to you – to predict compatible romantic partners. Facebook recently experimented with manipulating the emotional content of more than half a million of its users news feeds. Through altering the number of positive and negative posts users saw, based on a simple sentiment analysis algorithm, the psychological study examined how emotions could be spread through social media.[10] The study found that sentiment is in fact predictably contagious across social networks, where positive expressions beget the same, with the converse also true.[11]

9. Research shows that if you can capture the customer prior to the birth of a child, they are far more likely to purchase all of the various items they will need from you. See Charles Duhigg, "How Companies Know Your Secrets," *New York Times*, February 16, 2012. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

10. Vindu Goel, "Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry," *New York Times*, June 29, 2014. http://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html

11. Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," in *Proceedings of the National Academy of Sciences 111* (2014).

Predictive analytics has long played a role in military and intelligence applications. Cybernetics pioneer Norbert Weiner worked on predicting flight paths and evasive manoeuvres of German warplanes in World War II to assist in their targeting. Today, ArcGIS, a leading software platform for working with Geographic Information Systems (GIS), offers a Predictive Analysis Tools Add-in that provides drone operators the ability to query digital aerial photographs, imagery from satellites, digital pictures, or even scanned maps to build predictive targeting models from environmental data-sets, known parameters or historic sightings.[12] This tool enables the prediction of targets based on doctrine (established rules for identifying target locations), or when a clear doctrine is absent, using evidence based on Signals Intelligence (SIGINT) and the like to construct a doctrine based on locations the target has been known to frequent.[13] It is interesting to note that while the military has historically been the originator of technology transfer for commercial application – GPS technology or the Internet, for example – for ArcGIS it has become just another market segment for their spatial analysis and prediction products, alongside government agencies, real estate developers, and urban planning firms.

12. See http://solutions.arcgis.com/military/

These predictive processes take on qualities

13. For more on how mobile phone SIM cards are currently used by the U.S. Military for locating targets, see https://firstlook.org/theintercept/2014/02/10/the-nsas-secret-role/

of what I have described elsewhere as sentience.[14] The term "sentient" foregrounds the subjective relations that underlie technologies currently being promoted by high-tech marketing departments as "smart".[15] Sentience, derived from the Latin word "sentire", refers to the ability to feel or perceive subjectively, but does not necessarily include human faculties of reasoning or self-awareness. Which is to say, the possession of "sapience", meaning "to know", is not required for something to be sentient. Here it is important to differentiate between the act of sensing and that of having a sensation. Sensing, the thinking goes, is something animals, some plants, and some machines can do. It involves a sensing organ or device that enables the organic or inorganic system of which it is a part to actively respond to things happening around it. An organism or system may sense heat, light, sound, or the presence of rain, for example. Yet having a sensation or a feeling is something that goes beyond mere sensing, for it involves an internal state in which information about the environment is processed by that organism or system so that it comes to have a subjective character. "Qualia" is the philosophical term for this, which Daniel Dennett defines as "an unfamiliar term for something that could not be more familiar to each of us: the ways things seem to us".

14. Mark Shepard, *Sentient City: Ubiquitous Computing, Architecture, and the Future of Urban Space* (New York, Cambridge: Architectural League of New York; MIT Press, 2011).

15. See for example IBM's Smarter Planet initiative, http://www.ibm.com/smarterplanet/us/en/

The capacity for predictive analytics to process data in ways that lend it subjective qualities – this customer *seems* pregnant, this couple *seems* like a good match, this location *seems* like a good one for a new store, this caravan of pickup trucks *seems* like a good target – upends Cartesian distinctions between human and non-human beings.[16] Here questions of agency are foregrounded. If a predictive model developed by a hedge fund for high frequency trading crashes the market, who's to be held accountable? Who, or perhaps more precisely what, is responsible when a drone strike on a target identified using predictive analytics is made in error? Does the responsibility lie with the person who developed the predictive model? The agency that provided the data upon which the model was developed? The organisation that chose to apply that model to a specific situation? These questions are compounded when algorithms operate at levels of abstraction that exceed the ability of people who wrote them to understand the high level descriptors that indicate the likelihood of a future event occurring or not occurring. Here, non-human subjectivity becomes a black box of the first order.

16. René Descartes argued in *Meditations on First Philosophy* that sentience was an essential human capacity, whereas the behaviour of animals, for instance, could be accounted by purely physical processes involved in mere sensing.

When applied to geographies of information, the notion of sentience affords thinking of information systems and infrastructures as complex socio-technical entanglements of people, data, and space that come into being through a mutually constitutive process. These geographies need to be understood ontogenetically, as something continually brought into being through specific practices that alter the conditions under which space itself is (re) produced. Whereas predictive models indicate the probability of individual action, agency in this context is distributed across a range of human and non-human actors, and is understood more in collective terms. Consider for example the assemblage of drivers, passengers, smart phones, apps, networks, city streets, fare models, policies, regulations, investors, employees, and so on that constitute the urban car service Uber. When demand is high at a given location, a mechanism called surge-pricing kicks in, by which the standard fare for a ride is multiplied by a factor proportional to the degree of scarcity of available cars nearby. This inflated fare – which can often rise to at least eight times the standard fare – is intended to incentivise drivers to migrate in the direction of that scarcity. Here the spatial distribution of resources (available cars) is balanced through an economic incentive (surge-pricing) based on real-time analytics of fine-grained, geolocated demand (passengers).[17]

This has implications for how we think about not only the material basis of urban form but also the theatre of operations for practices that seek to engage urban life within these hybrid conditions. The notion of a gated community (or office tower) changes when access to physical space can be granted or denied in real-time based on what an algorithm anticipates an individual may or may not do. Predictive models alter how policies are formulated and decisions are made regarding the distribution of urban services and resources such as the locations of parks and hospitals, or the strategic pre-positioning of ambulances and police cars. New global connections between locales, cities and regions alike are created when passenger airlines dynamically determine flight routes based on emerging markets predicted by algorithms. When historical facts and probable futures converge in an algorithmic present, acting upon the spatial organisation and material disposition of these diverse conditions becomes a question of negotiating relations between people, data, code, and space of scales and durations.

Finally, we might address conditions of error within these new geographies. Within the field of predictive analytics, it is standard

17. It didn't take long for people to figure out how to out-smart the algorithm, however. If one waited five minutes or so, the system would balance itself and the fare price would return to normal. Uber itself publicly announced this way of hacking its surge-pricing system in response to strong reactions to what was widely perceived as price gouging.

practice for statisticians to conduct tests in order to determine whether or not a speculative hypothesis concerning observed phenomena of the world (or its inhabitants) can be supported. The results of such testing determine whether a particular set of results agrees reasonably (or does not agree) with the speculated hypothesis. In experimental science, the null hypothesis – the hypothesis to be tested – is generally a statement that there is no relationship between the value of a particular measured variable and that of an experimental prediction. The accuracy of such predictions is measured as a percentage of false positive (or false negative) results. A false positive is considered an error of the first order in statistical analysis (a type I error), where the error leads one to conclude that a supposed relationship between two entities exists when in fact it does not. It is the incorrect rejection of a true null hypothesis.

What will we make of geographies produced through false positives? One could look to the geographies of foreclosure emerging in the wake of the 2008 financial crisis for clues. As predictive analytics plays an increasing role in real-estate development – anticipating desirable neighbourhoods and markets – we will inevitably see new patterns of both density and vacancy emerge.[18] When algorithms that can predict the crime rate of a neighbourhood using Google Street View are added to the feature list, errors in Neighbourhood Home Price Indexes and Automated Valuation Models could lead to new cycles of real-estate booms and busts.[19] With the occasional predictive failure in audience-level targeting and hyper-local demand generation, we can look forward not only to inexplicable micro-islands of density, but also to ghost streets, buildings and individual homes cropping up in odd and unexpected places. Given the highly optimised, ever-more efficient and over-coded nature of these new geographies, it is perhaps within their unintended anomalies that we'll find the more interesting terrain for thought and action.

Urban media art is well positioned to critically investigate these ecologies of people, data, code, and space that constitute contemporary urban life. Visualising how algorithms do work in the world contributes to greater public literacy in increasingly data-driven environments, especially urban ones. Rendering the processes of predictive analytics in tangible and more accessible terms opens up these black boxes to greater public scrutiny, and raises new questions for urban policy and regulation. Showing how both human and non-human actors are capable of exhibiting

18. See for example http://www.smartzip.com/corp/technology/predictive-analytics/

19. See for example research by MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) as reported here: http://www.fastcolabs.com/3036677/this-algorithm-knows-your-neighborhood-better-than-you-do

sentient qualities – such as subjective perception and feeling – upends traditional notions about the relations between a city and its citizens. Highlighting error in the apps, devices, systems and infrastructures that operate within, and upon, urban space underscores their frailty, and ultimately, their humanity. Taken together, the diverse communities of practice that constitute urban media art offer not simply the means to represent contemporary urban space, but more importantly to actively engage it in ways that can effect change across a range of scales and durations.

*Originally published in Taraneh Meshkani and Ali Fard, *New Geographies 07: Geographies of Information* (Cambridge: Harvard University Graduate School of Design, 2015), 133.