PERSPECTIVE

THE USES OF BIG DATA IN CITIES

Luís M.A. Bettencourt
Santa Fe Institute, Santa Fe, New Mexico

Abstract

There is much enthusiasm currently about the possibilities created by new and more extensive sources of data to better understand and manage cities. Here, I explore how big data can be useful in urban planning by formalizing the planning process as a general computational problem. I show that, under general conditions, new sources of data coordinated with urban policy can be applied following fundamental principles of engineering to achieve new solutions to important age-old urban problems. I also show that comprehensive urban planning is computationally intractable (i.e., practically impossible) in large cities, regardless of the amounts of data available. This dilemma between the need for planning and coordination and its impossibility in detail is resolved by the recognition that cities are first and foremost self-organizing social networks embedded in space and enabled by urban infrastructure and services. As such, the primary role of big data in cities is to facilitate information flows and mechanisms of learning and coordination by heterogeneous individuals. However, processes of self-organization in cities, as well as of service improvement and expansion, must rely on general principles that enforce necessary conditions for cities to operate and evolve. Such ideas are the core of a developing scientific theory of cities, which is itself enabled by the growing availability of quantitative data on thousands of cities worldwide, across different geographies and levels of development. These three uses of data and information technologies in cities constitute then the necessary pillars for more successful urban policy and management that encourages, and does not stifle, the fundamental role of cities as engines of development and innovation in human societies.

New Opportunities for the Use of Big Data in Cites

> How does one measure a city? By the buildings that fill its skyline? By the efficiency of its rapid transit? Or, perhaps, by what Jane Jacobs called the "sidewalk ballet" of a busy street? Certainly these are the memorable hallmarks of any modern city or metropolitan area. But a city's true measure goes beyond human-made structures and lies deeper than daily routine. Rather, cities and metro areas are defined by the quality of the ideas they generate, the innovations they spur, and the opportunities they create for the people living within and outside the city limits.
>
> —Judith Rodin[1]

The rise of information and communication technologies (ICTs)[2,3] and the spread of urbanization[4] are arguably the two most important global trends at play across the world today. Both are unprecedented in their scope and magnitude in history, and both will likely change the way we live irreversibly. If current trends continue, we may reasonably expect that the vast majority of people everywhere in the world will live in urban environments within just a few decades and that

BIG DATA MARCH 2014 • DOI: 10.1089/big.2013.0042

information technologies will be part of their daily lives, embedded in their dwellings, communications, transportation, and other urban services. It is therefore an obvious opportunity to use such technologies to better understand urbanism as a way of life[5,6] and to improve and attempt to resolve the many challenges that urban development entails in both developed and developing cities.[4,7]

Despite its general appeal, the fundamental opportunities and challenges of using big data in cities have, in my opinion, not been sufficiently formalized. In particular, the necessary conditions for the general strategic application of big data in cities need to be spelled out and their limitations must also be, as much as possible, anticipated and clarified. To address these questions in light of the current interdisciplinary knowledge of cities is the main objective of this perspective.*

Before I start, it is important to emphasize that the use of quantitative data to better understand urban problems and to guide their solutions is not at all new. In the United States—and in New York City in particular—we have been building detailed statistical pictures of urban issues for over a century. Jacob Riis's influential *How the Other Half Lives*[11] derived much of its persuasive power from the use of statistics, for example, by giving numbers of deaths to the person in tenements in New York City. The NYC-RAND Institute in the 1970s[12] used detailed urban statistics, modeling, and computation developed for wartime and corporate management to determine resource allocation, especially for New York City's fire department.[13] Thus, today's *Smart Cities* movement[14–28] needs to be placed in perspective: Are the achievements made possible by modern data and information technologies fundamentally different from what was possible in the past? Or are they the extensions of old ideas and procedures, although with greater scope and precision?

"ARE THE ACHIEVEMENTS MADE POSSIBLE BY MODERN DATA AND INFORMATION TECHNOLOGIES FUNDAMENTALLY DIFFERENT FROM WHAT WAS POSSIBLE IN THE PAST?"

To answer these questions I formalize the use of data in urban policy and management in light of the conceptual frameworks of engineering.[29] I show, in the next section, that this allows us to identify the necessary conditions for the effective use of data in policies that address a large array of urban issues. In this way, I will demonstrate that, in principle at least, modern ICTs are opening entirely new windows of opportunity for the application of engineering solutions to cities.

But the problems of cities are always primarily about people. Social and economic issues in cities define what planners call "wicked problems,"[30] a term that has gained currency in policy analysis well beyond the urban context. These are the kind of issues that are not expected to yield to engineering solutions, for specific reasons[30] that break the assumptions of feedback control theory.[29] In section 3, I revisit the issue of wicked problems in light of computational complexity theory[31] to make the formal argument that detailed urban planning is computationally intractable. This means that solutions requiring the knowledge and forecast of chains of detailed behaviors in cities have the fundamental property that they become *practically impossible* in large cities, regardless of the size of data available. This clarifies the central dilemma of urban planning and policy: planning is clearly necessary to address long-term issues that span the city, for example, in terms of services and infrastructure, and yet the effects of such plans are impossible to evaluate *a priori* in detail.

In section 4, I resolve this dilemma. The key is the nature of self-organization of social and economic life in cities and the development of a general quantitative understanding of how such processes operate in cities as vast networks, spanning large spatial and temporal scales. Theory recognizes that most individual details are irrelevant to describe complex systems as a whole, while identifying crucial general dynamics.[32] For example, a city exists and functions even as people change their place of residence, jobs, and friendships. The development of this kind of urban theory too follows from increasing data availability in many cities around the world, both in terms of observations and, wherever possible, from experiments.

I then discuss how the increasing integration of limited-scope engineering solutions with the dynamics of social self-organization in cities, articulated by the long and large views afforded by urban theory, provides the appropriate context to understand and manage cities, while allowing them to play their primary function in human societies of continuing to evolve socially and economically in open-ended ways.

The Nature of Big Data Solutions in Cities

> It is a profoundly erroneous truism [...] that we should cultivate the habit of thinking what we are

*In this article I do not discuss issues of ethics or privacy related to big data, nor of the political or corporate potential pitfalls of using big data to manage cities. This is not because these issues are not important but because here I wish to focus on the fundamental possibilities to manage cities based on data, without such complications. It can probably be said that any of these considerations will work to limit the scope and effectiveness of the use of data to address urban issues and will, as a consequence, make the case for its use less compelling. Several authors have recently written eloquently about several of these issues.[8–10] doing. The precise opposite is the case. Civilization advances by extending the number of important operations, which we can perform without thinking about them.

—Alfred Whitehead[33]

Relatively simple-minded solutions, enabled by precise measurements and prompt responses, can sometimes operate wonders even in seemingly very complex systems where traditional policies or technologies have failed in the past. An example, not immediately related to managing cities, illustrates this point: self-driving cars. I have recently attended a workshop at the Kavli Institute for Theoretical Physics in Santa Barbara, CA, as part of a group of interdisciplinary researchers discussing issues of artificial intelligence, algorithms, and their relation to neuroscience. In a particularly memorable lecture, Prof. Richard Murray from Caltech[†] explained how his team engineered a self-driving vehicle that completed DARPA's *Grand Challenge* in 2005[34] and was now deploying it on the road in cities.[‡] Richard described the array of diverse sensors for determining surrounding objects and their speed, for geolocating the vehicle, recognizing the road and its boundaries, and planning its path. I interjected, asking whether the vehicle learned from its experience and, to my surprise, the response was no: it was all "hardwired" by the team at Caltech, using principles of feedback control theory. I protested that surely the hardest problem faced by a self-driving car was not planning its path, recognizing the road, or measuring surrounding traffic—but all the stupid things that people do on the road: "Doesn't the car need a model of human behavior? That is surely the hard part!"

The answer, surprisingly, is no, not at all! The reason is that while people surely do reckless things behind the wheel, they do so, from the car's perspective, at a truly glacial pace: While humans behave unpredictably on the timescale of seconds, the car's electronics and actuators respond in milliseconds or faster. In practice, this allows the car more than enough time to take very basic "hardwired" evasive actions, such as breaking or getting out of the way. In fact, presently self-driving cars have a safety record that exceeds that of humans.[35,36]

The lesson from this story is that relatively simple solutions that require no great intelligence can, under specific conditions, solve very hard problems. The key is fast and precise enough measurement and adequate simple reactions. In other words, sometimes you don't need to be very smart if you are fast enough.

This is the logic of modern engineering, and more precisely of feedback control theory.[29] If we know the desired operating point for a system (e.g., avoid collision), and we have the means to operate on the system, as we observe its state change via feedback loops, we can under general conditions turn it into a simple problem. The crucial conditions are to be able to measure and recognize potential problems just as they start to arise (car approaching) and act to make the necessary corrections (break or get out of the way, thereby avoiding collision). The crucial issue is that of temporal scales; every system has intrinsic timescales at which problems develop—a few seconds in the case of a car on the road. Cycles of measurement and reaction that avoid such complex problems by simple means must act well within this window of opportunity.

We now see more clearly why a new generation of bigger data may offer radically novel solutions to difficult problems. Modern electronics are now so fast in comparison to most physical, biological, and social phenomena that myriads of important policy problems are falling within this window of opportunity. In such circumstances, models of system response can be very simple and crude (they can typically be linearized).[29] Thus, the engineering approach conveniently bypasses the complexities that always arise in these systems at longer temporal or larger spatial scales, such as the need to develop models of human behavior. In this way, a difficult and important problem can be solved essentially *without theory*;[37] this is the (potential) miracle of big data in cities.

"IN THIS WAY, A DIFFICULT AND IMPORTANT PROBLEM CAN BE SOLVED ESSENTIALLY WITHOUT THEORY;[37] THIS IS THE (POTENTIAL) MIRACLE OF BIG DATA IN CITIES."

Many examples of urban management and policy in cities that use data successfully have this flavor, regardless of being implemented by people and organizations or by computer algorithms. Table 1 presents a summary of important urban issues, where I attempted to roughly characterize their typical temporal and spatial scales and the nature of their operating points, or outcomes.

For example, think of urban transportation systems, such as a bus network, in these terms.[38] Buses should carry passengers who wait a few minutes to be transported over a few

[†]This paragraph is based on my personal recollection of Richard Murray's lecture. Any inaccuracies or embellishments are mine, not the speaker's.
[‡]Self-driving cars or autonomous vehicles (most famously developed by Sebastian Thrun's team from Stanford University, now at *Google*) are no longer news in urban environments. *Google* deploys a dozen cars on a typical day in the Bay Area.[35]

Table 1. Urban Issues, Their Temporal and Spatial Scales, and the Character of Their Associated Metrics

Problem	*Timescale*	*Spatial scale*	*Outcome metric*
Transportation (buses, subway)	Minutes	Meters	Simple
Fire	Minutes	Meters	Simple
Epidemics (HIV, influenza)	Years, days	Citywide	Simple
Chronic diseases	Decades	Citywide	Simple
Sanitation	Years	Citywide	Simple
Crime	Minutes	Meters	Simple
Infrastructure (roads, pipes, cables)	Days	Meters	Simple
Traffic	Minutes	Meters to km	Simple
Trash collection	Days	Meters	Simple
Education	Decades	Citywide	Complex
Economic development	Decades	Citywide	Complex
Employment	Years	Citywide	Complex
Poverty	Decades	Neighborhood	Complex
Energy and sustainability	Years	Citywide	Complex
Public housing	Years to decades	Neighborhood	Complex

Outcome metrics are to be interpreted on the timescale in the table. On longer timescales, socioeconomic issues, characterized by long times, become part of every problem.

kilometers. Measuring the time in between buses at each stop, possibly together with the number of passengers waiting, gives the planner the basis for a feedback control solution: Communicate with buses to enforce desired standards of service, quickly place more or fewer units in service where these parameters start to deviate from the ideal metrics, and the quality of service as measured by per person waiting times will improve. This type of strategy can be operated intuitively by human dispatchers but possibly can also be implemented automatically by an ICT algorithm[38] with access to the necessary measurements and actions. Feedback control theory provides the framework for the development and optimization of any of these solutions.

Similar procedures could be devised for water or power supply management and for their integration with public transportation systems.[39,40] Traffic management, trash collection, and infrastructure maintenance (building, roads, pipes, etc.) could also generally be thought of, and integrated together, in analogous ways.

There is no reason in principle why certain stubbornly difficult social problems, such as crime, may not be responsive to patrolling and other law enforcement strategies that follow similar principles. Crime is local and its outcomes are clear enough. Exciting new experiments in several cities, notably in Los Angeles,[41,42] suggest that much.

Recent insights by innovative corporations[43–46] and policy strategists[47] have correctly anticipated that flows of information underpin such coordinated solutions. Thus, progress in these problems is fundamentally an ICT problem, enabled by simple actions or policies that nudge the states of systems toward optimal performance.

But other urban issues, especially those that are primarily social or economic, acquire, as seen through this lens, a different character because their operating points are not well defined or because their dynamics are rather diffuse and play out over large temporal or spatial scales (Table 1). Thus, it has remained difficult to create engineering solutions to problems of education, public housing,[§] economic development, poverty, or sustainability at the city level.

Public health issues often sit at an intermediate level of complexity. Analogously to crime, contagious diseases are often characterized by simple metrics and by local processes of social contact between individuals. Thus, containment or eradication of cholera or of contagious diseases for which vaccination exists represents some of the greatest successes of urban policy.** But conditions that play out over longer times[††] and possibly have more complex and diffuse social causation, such as chronic diseases or even HIV/AIDS, have proven far more difficult.

Thus, the simplicity of performance metrics expressed as objective quantitative quantities and knowledge of their proximate causes in space and time are the crucial conditions for successful engineering-inspired solutions. The important point to always bear in mind, however, is that these quantities are relative to the properties of the controlling system—the policy maker or the algorithm—such as their response times, so that with technological and scientific evolution we may hope to fulfill Alfred Whitehead's maxim quoted at the beginning of this section, of achieving progress through the increasing automation of solutions to once intractable problems.[‡‡]

[§]Notably, public housing policy in places such as Hong Kong[48] or Singapore[49] has fared uncharacteristically well compared with Western cities. These successes are likely because of a combination of an adaptable design, where apartments can be expanded as households become more prosperous, careful matching between housing characteristics and socioeconomic household capacities, and close monitoring of housing problems and prompt response, in terms of conflict resolution, maintenance, and repair.

**Dealing with the problem of cholera, by John Snow in London in the 1840s–50s, constitutes one of the first examples of a statistical approach to public health, though one thoroughly based on "shoe-leather epidemiology."[50]

[††]An exception to this argument is influenza. It is a "fast" disease that has defied public health response times. Its symptoms are generic and it typically is a large-scale problem, affecting 20–30% of a population over large spatial regions.

[‡‡]The arrow of progress made possible by engineering and automation does not always run in the same direction. The self-driving-car solution that motivated this section works only as long as most cars are driven by people. Once most cars are self-driven and share the same fast response times, new instabilities will develop. A preview of such problems can be seen in stock markets today, where it has been estimated that maybe over 80% of trades are made by algorithms on ultrafast timescales.[51] Often such algorithms act on news written by other ultrafast algorithms.

The Planner's Problem

> In order to describe a wicked-problem in sufficient detail, one has to develop an exhaustive inventory of all conceivable solutions ahead of time. The reason is that every question asking for additional information depends upon the understanding of the problem—and its resolution—at that time. [...] Therefore, in order to anticipate all questions (in order to anticipate all information required for resolution ahead of time), knowledge of all conceivable solutions is required.
>
> —Rittel and Webber[30]

The scenarios developed in the previous section paint a well-defined path for the use of data in cities, at odds with many decades of urban theory and practice. While many physical and infrastructural aspects of cities appear at first sight manageable through engineering practices, many social and economic issues are in fact "wicked problems."[30] Over the long run, the social and infrastructural aspects of the city are always entangled.

Wicked problems were originally defined by two urban planners, Rittel and Webber, in 1973.[30] The essential character of wicked problems is that they cannot be solved in practice by a (central) planner. I shall reformulate some of their arguments in modern form in what I like to call the "planner's problem." The planner's problem has two distinct facets: (i) the knowledge problem and (ii) the calculation problem.

The knowledge problem refers to the information (ultimately as data) that a planner would need to map and understand the current state of the system—the city in our case. While still implausible, it is not impossible to conceive ICTs that would give a planner sitting in a "situation room" access to detailed information about every aspect of the infrastructure, services, and social lives in a city. Privacy concerns aside, it is conceivable that the lives and physical infrastructure of a large city could be adequately sensed in several million places at fine temporal rates, producing large but manageable rates of information flow by current technology standards.[§§] In this way, the knowledge problem is not a showstopper.

"GIVEN THESE CAVEATS, I HAVE SHOWN THAT PLANNING THE CITY IN DETAIL BECOMES COMPUTATIONALLY IMPOSSIBLE IN ALL BUT THE SMALLEST TOWNS."

The calculation problem, however, alluded to in the quote at the beginning of this section, refers to the computational complexity to perform the actual task of planning in terms of the number of steps necessary to evaluate all possible scenarios and choose the best possible course of action. This problem is analogous to playing a complex game, such as chess, but with millions of pieces, each following different rules of interaction with others, against millions of opponents, on a vast board. It will therefore not surprise the reader that the exhaustive approach of evaluating each possible scenario in a city is impractical as it involves the consideration of impossibly large spaces of possibilities.

I formalize this statement in the form of a theorem and provide mathematical details and the sketch of a proof in the Appendix. The identification of the best detailed urban plan as a result of listing and evaluating all the the possibilities configurations of a city would require a computation involving $O[P(N)]$ steps in a city with population N, where

$$P(N) \simeq P_0 e^{N^{1+\delta} \ln N},$$

where $\delta \simeq 1/6$ and P_0 is a constant, independent of N. For a city of a million people, this is $P(N) = P_0 10^{6\times10^6}$, a truly astronomical number much, much larger than all the atoms in the universe. Thus, the planner's problem is impossible to solve in practice.

I want to emphasize that the sketch of a proof given in the Appendix relies on the formulation of the planner's problem as the choice of the best spatial and social configuration for the city that maximizes a set of urban metrics. A more limited conceptualization of the problem will likely have lower computational complexity, but may not guarantee finding the best plan. In such a case, planning must rely on heuristics and cannot guarantee absolute optimality.

Given these caveats, I have shown that planning the city in detail becomes computationally impossible in all but the smallest towns. This indicates that the use of complex models[39,52] in the detailed planning and management of cities has its limits and cannot be exhaustively mapped and solved in general, regardless of how much data a planner may acquire. How then should we think of planning under these limitations, and what becomes the role of bigger and better data in cities?

[§§]For the largest city in the world, Tokyo, with 35 million inhabitants, we could imagine sensing each individual and a corresponding number of places, including different infrastructure and services using, say, 100 million data streams, sampled at 1 byte each once a second (a faster rate may be desirable for some applications, slower in others). This results in information flows of the order of 10^9 bits per second (Gbps) that are already handled by gigabit Internet and current-generation computing. So, while still daunting, these numbers can already be handled using current technologies.

Self-Organization, Information, and the Role of a Science of Cities

> Provided that some groups on earth continue either muddling or revolutionizing themselves into periods of economic development, we can be absolutely sure of a few things about future cities. The cities will not be smaller, simpler or more specialized as cities of today. Rather, they will be more intricate, comprehensive, diversified and larger than today's and will have even more complicated jumbles of old and new things as ours do.
>
> —Jane Jacobs[53]

Part of the answer to the planner's problem comes from economics, where the social (central) planner faces a similar, if more limited, predicament of organizing economic markets. The detailed information necessary to run a city in terms of planning the lives of people and organizations is, of course, best known to those agents themselves. Even more importantly, the motivation and capacity to make good decisions and act on this information resides (privately) with the very same agents. Thus, cities are self-organizing and depend crucially not on the detailed instructions from the planner's situation room, but rather on the integration and coordination of myriads of individual decisions, made possible by suitable information flows.***

In this sense, cities have always been the primary creators and users of ICTs, from the daily newspaper, to postal mail, to the telegraph, or the (cellular) telephone.[56] It should then come as no surprise that new ICTs will primarily act to further enable cities rather than compete with them. In fact, the problem that ICTs address in specific niches is the general problem solved by cities[57]: That is, the integration and coordination of people and organizations through information sharing in vast social networks that can confer benefits to individual agents and to society as a whole.

This then brings us to another perspective on the use of big data in cities, radically different from the engineering framework of the previous sections. A vital function of ICTs is in enabling new and better ways for people to be social. In this respect, big data and associated technologies play only a supportive role to social dynamics, not a prescriptive one. Data and technologies, then, do not create or solve urban problems so much as they enable people and social organizations to address them better. This role of ICTs has been discussed in the literature on smart cities.[19,21,25,26] Especially important is the potentially more rapid access to civic and economic participation of poor or excluded populations, particularly in developing cities and nations.[58]

"DATA AND TECHNOLOGIES, THEN, DO NOT CREATE OR SOLVE URBAN PROBLEMS SO MUCH AS THEY ENABLE PEOPLE AND SOCIAL ORGANIZATIONS TO ADDRESS THEM BETTER."

The second part of the answer stems from the fact that bottom-up self-organization is clearly not enough for a well-functioning city. Many developing cities—think of Mumbai[59]—have self-organization in spades. Similarly, a flash-mob is as much the product of self-organization as a new art form or a new business. For cities to work, general properties and constraints that define the city on larger scales must be at play. Such features, common to all cities, are the focus of urban theory.

Scientific theory is not about the details of a particular practical problem; it is about general principles that apply across many problem instances. Given how different cities are from each other and how much they change over time, to what extent can a scientific theory of cities be possible or useful?

To illustrate this point, consider another great triumph of science and engineering: the Moon shot. It is not completely inconceivable that the Apollo 11 mission could have been pulled off in the absence of knowledge of the laws of motion and gravity. One could imagine controlling rocket thrusters using feedback by judging whether the spacecraft was leaving Earth, approaching the Moon at the right speed not to crash, and so on. But of course this is not at all what happened: such solutions would be very difficult in terms of engineering and computation at the time, especially because of the fast timescales and large energies necessary for liftoff. Instead, we launch rockets based on their general properties as prescribed by physical theory, the mass of the Earth and of the spacecraft, the rotation of the planet, the speed necessary at liftoff, and so forth. This set of calculations gives a general strategy for launching *any* spacecraft from Earth and placing it in desired orbital positions.

What follows this general solution is a detailed, but much more limited, use of engineering principles and self-organization to fine-tune solutions to the details of the problem, such as the conditions for Moon landing given

***The original argument for self-organization in the context of economics was made, for example, by Friedrich Hayek in the essay "The Use of Knowledge in Society,"[54] a title that the present article echoes. The central argument is that economies are primarily about information and that the price system of economics is one (good) way of achieving the necessary self-organization in terms of the allocations of goods and services, labor, and capital. This point was also emphasized by Paul Krugman in his 1992 Ohlin lectures.[55]

the particular spacecraft's speed at that moment, its distance to the surface, the human operators' past actions, and the like.

Thus, theory can enormously reduce the size of an engineering problem and in many cases is essential to make such solutions possible. In this way, urban theory sets the general parameters for *any* city to work as a whole as a bottom-up, self-organizing process mediated by social networks with certain general properties and enabled by infrastructure and services that follow general performance metrics.[57] The details of these processes must then be measured and managed, as appropriate, in every situation.

The present state of urban theory is fast evolving and is starting to approach the status of an interdisciplinary science. New, bigger data that allows for the comparative quantitative analysis of thousands of cities worldwide has been the key in determining the general properties of cities and their variation due to both general conditions of development and local histories.[57,60,61] We now have a quantitative framework that connects the general structure of cities as vast social networks, to their general properties of infrastructure networks and land use.[57] In this way, we are uncovering what kind of complex system cities really are[6,57,62] and the general conditions in terms of infrastructure, energy and resource use, and so on, that allow cities to function optimally as distinct open-ended social systems capable of socioeconomic growth and innovation in human societies.[†††]

Discussion

> The last great technological advancement that reshaped cities was the automobile (some might argue it was the elevator). In both cases, these technologies reshaped the physical aspects of living in cities—how far a person could travel or how high a building could climb. But the fundamentals of how cities worked remained the same. What's different about the information age that has been ushered in by personal computers, mobile phones and the Internet is its ability to reshape the social organization of cities and empower everyday citizens with the knowledge and tools to actively participate in the policy, planning and management of cities.
>
> —Christian Madera[65]

Developed cities today are social and technical complex systems characterized by historically unprecedented levels of diversity and temporal and functional integration. This growing individual specialization and interdependence makes large cities extremely diverse and crucially relies on fine temporal and spatial integration and on faster and more reliable information flows. These informational processes lie at the core of what makes cities the economic and cultural engines of all human societies.

A city of 35 million people, like present-day Tokyo, is made possible by extremely efficient transportation, reliable energy and water supply, and good social services, including a very low level of conflict. Many developing urban areas in the world today will have to replicate and improve upon these metrics if they are to fulfill their promise to become world cities and enable the pace of human development expected by millions of their inhabitants.

"THESE INFORMATIONAL PROCESSES LIE AT THE CORE OF WHAT MAKES CITIES THE ECONOMIC AND CULTURAL ENGINES OF ALL HUMAN SOCIETIES."

I have showed here that the increasing separation of timescales between ICTs and the natural pace of human behavior will enable many fundamentally new engineered solutions that may solve important age-old urban problems. Thus, new sensing, communications, and computational technologies will play a vital role in enabling new urban-scale efficiencies and city socioeconomic growth. In this sense, we should expect that with urban development, services and infrastructure should become increasingly engineered, invisible, predictable, and automatic. How resilience to infrastructure and service failures may be maintained in these circumstances is a crucial open problem. Important as they are, such solutions should always be understood to be short-term and to have a limited scope, as they must continue to adapt and enable socioeconomic development over the long run.

The world's most vibrant and attractive cities are not usually the same places where the buses run impeccably on time. While improvements in infrastructure and urban services are absolutely necessary for cities to function better, they are not the fundamental sources of social development or economic growth. These more fundamental functions of cities rely on processes of social self-organization and on the fulfillment of general conditions that allow cities to operate effectively as multilevel open-ended evolving networks. It is, curiously, this more fundamental character of cities—at play from the individual to the city as a whole and from seconds to centuries—that creates and solves the planner's problem.

[†††]While recent progress in urban theory has allowed us to better understand cities as multilevel interacting networks and as processes of self-organization across scales, more work needs to be done regarding heterogeneous aspects of the city (inequality of incomes, variations across neighborhoods[63]) and the evolution of cities over time, including the processes of social development and economic growth.[53,64]

In this light, the key to making cities even smarter, through new uses of ICTs, is not very different from what it has always been. Big data creates more opportunities, not less, for cities to accentuate their urban character: The primary uses of big data in cities must then be to continue to enable the creation of new knowledge by more people, not replace it.

Acknowledgments

I thank the participants of the workshop "How Far Can Big Data Take Us Towards Understanding Cities?" at the Santa Fe Institute, September 19–21, 2013, where an early version of this article was presented. This research was partially supported by the Rockefeller Foundation, the James S. McDonnell Foundation (Grant No. 220020195), the John Templeton Foundation (Grant No. 15705), the Army Research Office (Grant No. W911NF1210097), the Bill and Melinda Gates Foundation (Grant OPP1076282), and a gift from the Bryan J. and June B. Zwan Foundation.

References

1. Rodin J in the foreword to Katz B, Bradley J. The Metropolitan Revolution: How Cities and Metros Are Fixing Our Broken Politics and Fragile Economy. Washington, DC: Brookings Institution Press, 2013.
2. Hilbert M, López P. The world's technological capacity to store, communicate, and compute information. Science 2011; 332:60–65.
3. Mayer-Schonberger V, Cukier K. Big Data: A Revolution That Will Transform How We Live, Work, and Think. New York, NY: Eamon Dolan/Houghton Mifflin Harcourt, 2013.
4. UN-Habitat. Planning Sustainable Cities—Global Report on Human Settlements 2009. Available online at www.unhabitat.org/content.asp?typeid=19&catid=555&cid=5607
5. Wirth L. Urbanism as a way of life. Am J Sociol 1938; 44:1–24.
6. Bettencourt LMA. The kind of problem a city is. In: Offenhuber D, Ratti C (eds.). Die Stadt Entschlusseln: Wie Echtzeitdaten Den Urbanismus Verandern: Wie Echtzeitdaten den Urbanismus verändern. Heidelberg: Birkhauser, 2013: 175–187. Available online (in English) at www.santafe.edu/media/workingpapers/13-03-008.pdf
7. UN-Habitat. The Challenge of Slums: Global Report on Human Settlements 2003. Available online at www.unhabitat.org/pmss/listItemDetails.aspx?publicationID=1156
8. Polonetsky J, Tene O. Privacy and big data: making ends meet. 66 Stan. L. Rev. Online 25. Available online at www.stanfordlawreview.org/online/privacy-and-big-data/privacy-and-big-data
9. boyd d, Crawford K. Critical Questions for Big Data. Inf Commun Soc 2012; 15:662–679.
10. McKensey Global Institute. 2011. Big data: the next frontier for innovation, competition, and productivity. Available online at www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
11. Riis J. How the Other Half Lives: Studies Among the Tenements of New York. New York, NY: Kessinger Publishing, 2004. Originally published in 1890. Available online at www.authentichistory.com/1898-1913/2-progressivism/2-riis/index.html
12. Szanton PL. Analysis and urban government: experience of the New York City-Rand Institute. Policy Sci 1972; 3:153–161.
13. Flood J. The Fires: How a Computer Formula, Big Ideas, and the Best of Intentions Burned Down New York City and Determined the Future of Cities. New York, NY: Riverhead Books, 2011.
14. Gibson DV, Kozmetsky G, Smilor RW (eds.). The Technopolis Phenomenon: Smart Cities, Fast Systems, Global Networks. New York, NY: Rowman & Littlefield, 1992.
15. Castells M. The Informational City: Information Technology, Economic Restructuring, and the Urban-Regional Process. Oxford: Blackwell, 1989.
16. Batty M. The computable city. In: Fourth International Conference on Computers in Urban Planning and Urban Management, Melbourne, Australia, July 11–14, 1995. Available online at www.acturban.org/biennial/doc_planners/computable_city.htm
17. Graham S (ed.). The Cybercities Reader. London: Routlege, 2004.
18. Mitchell WJ. Me++: The Cyborg Self and the Networked City. Cambridge, MA: The MIT Press, 2004.
19. Partridge H. 2004. Developing a human perspective to the digital divide in the Smart City. Available online at http://eprints.qut.edu.au/1299/1/partridge.h.2.paper.pdf
20. Giffinger R, et al. Smart Cities Ranking of European Medium-Sized Cities. Vienna: Centre of Regional Science, Vienna University of Technology, 2007. Available online at www.smart-cities.eu/
21. Shapiro JM. Smart cities: quality of life, productivity, and the growth effects of human capital. Rev Econ Stat 2008; 88:324–335.
22. Lee SH, Han JH, Leem YT, Yigitcanlar T. Towards ubiquitous city: concept, planning, and experiences in the Republic of Korea. In: Yigitcanlar T, Velibeyoglu K, Baum S (eds.). Knowledge-Based Urban Development: Planning and Applications in the Information Era. Hershey: Information Science Reference, 2008, pp. 148–170.
23. Caragliu A, Del Bo C, Nijkamp P. Smart Cities in Europe. Research Memoranda 0048. Amsterdam: Business Administration and Econometrics, Faculty of Economics, VU University Amsterdam, 2009.
24. Harrison C, Eckman B, Hamilton R, Hartswick P, Kalagnanam J, Paraszczak J, Williams P. Foundations for smarter cities. IBM J Res Dev 2010; 54:1–16.

25. Taewoo N, Pardo T. Conceptualizing smart city with dimensions of technology, people, and institutions. In: The Proceedings of the 12th Annual International Conference on Digital Government Research, 2011. Available online at http://dl.acm.org/citation.cfm?id=2037602
26. Batty M, Axhausen KW, Giannotti F, Pozdnoukhov A, Bazzani A, Wachowicz M, Ouzounis G, Portugali Y. Smart cities of the future. Eur Phys J Spec Top 2012; 214:481–518.
27. Roche S, Nabian N, Kloeckl K, Ratti C. Are "smart cities" smart enough? In: Global Geospatial Conference, Global Spatial Data Infrastructure Association, 2012. Available online at www.gsdi.org/gsdiconf/gsdi13/papers/182.pdf
28. Papa R, Gargiulo C, Galderisi A. Towards an urban planner's perspective on smart city. TeMA 2013; 1:5–17.
29. Åström KJ, Murray RM. Feedback Systems: An Introduction for Scientists and Engineers. Princeton, NJ: Princeton University Press, 2008. Available online at www.cds.caltech.edu/~murray/amwiki/index.php/Main_Page
30. Rittel HWJ, Webber MM. Dilemmas in a general theory of planning. Policy Sci 1973; 4:155–169.
31. More C, Mertens S. The Nature of Computation. New York, NY: Oxford University Press, 2011.
32. Anderson PW. More is different. Science 1972; 177:393–396.
33. Whitehead AN. An Introduction to Mathematics. London: Williams & Norgate, 1911.
34. Team Caltech at the DARPA Grand Challenges. http://archive.darpa.mil/grandchallenge/TechPapers/Cal_Tech.pdf
35. Google driverless car. http://en.wikipedia.org/wiki/Google_driverless_car
36. Moore MM, Lu B. 2011. Autonomous vehicles for personal transport: a technology assessment. Available online at http://ssrn.com/abstract=1865047 or http://dx.doi.org/10.2139/ssrn.1865047
37. Anderson C. 2008. The end of theory: the data deluge makes the scientific method obsolete. Wired Magazine 16.07. Available online at www.wired.com/science/discoveries/magazine/16-07/pb_theory
38. Ren JSJ, Wang W, Liao SS. Optimal control theory in intelligent transportation systems research—a review. arXiv:1304.3778 [cs.SY].
39. Harrison C. 2013. Roads to smarter cities. Available upon request from the author.
40. MASDAR City. Available online at www.masdarcity.ae/en/
41. Short MB, D'Orsogna MR, Brantingham PJ, Tita GE. Measuring and modeling repeat and near-repeat burglary effects. J Quant Criminol 2009; 25:325–339.
42. Rubin J. Stopping crime before it starts. Los Angeles Times, August 21, 2010. Available online at http://articles.latimes.com/2010/aug/21/local/la-me-predictcrime-20100427-1
43. IBM Global Business Services. A Vision of Smarter Cities: How Cities Can Lead the Way into a Prosperous and Sustainable Future. Somers, NY: IBM Institute for Business Value, 2009.
44. Cisco's smart city framework. 2012. Available online at www.cisco.com/web/about/ac79/docs/ps/motm/Smart-City-Framework.pdf
45. Siemens' smart city in Vienna: model project Aspern. 2013. Available online at www.siemens.com/innovation/en/news/2013/e_inno_1319_1.htm
46. Smart Cities Council. Available online at http://smartcitiescouncil.com
47. European Union's Report of the Meeting of Advisory Group. 2011. ICT Infrastructure for energy-efficient buildings and neighborhoods for carbon-neutral cities. Available online at http://ec.europa.eu/information_society/activities/sustainable_growth/docs/smart-cities/smart-cities-adv-group_report.pdf
48. Dwyer DJ (ed.). Asian Urbanization, a Hong Kong Casebook. Hong Kong: Hong Kong University Press, 1971.
49. Fernandez W. Our Homes: 50 Years of Housing a Nation. Singapore: Housing and Development Board, 2011.
50. Johnson SB. The Ghost Map: The Story of London's Most Terrifying Epidemic—And How It Changed Science, Cities and the Modern World. New York, NY: Riverhead Books, 2007.
51. Johnson N, Zhao G, Hunsader E, Qi H, Johnson N, Meng J, Tivnan B. Abrupt rise of new machine ecology beyond human response time. Nat Sci Rep 2013; 3:Article no. 2627.
52. Portugali Y. Complexity, Cognition and the City. Heidelberg: Springer, 2011.
53. Jacobs J. The Economy of Cities. New York, NY: Vintage, 1970.
54. Hayek F. The use of knowledge in society. Am Econ Rev 1945; 35:519–530.
55. Krugman P. Development, Geography, and Economic Theory (Ohlin Lectures). Cambridge, MA: MIT Press, 1997.
56. Hall P. Cities in Civilization. New York, NY: Pantheon Books, 1998.
57. Bettencourt LMA. The origin of scaling in cities. Science 2013; 340:1438–1441.
58. Andris C, Bettencourt LMA. Development, information and social connectivity in Côte d'Ivoire, Santa Fe Institute Working Paper. Paper No. 13-06-023. Available online at www.santafe.edu/media/workingpapers/13-06-023.pdf
59. Mehta S. Maximum City: Bombay Lost and Found. New York, NY: Vintage, 2005.
60. Bettencourt LMA, Lobo J, Helbing D, Kühnert C, West GB. Growth, innovation, scaling, and the pace of life in cities. Proc Natl Acad Sci USA 2007; 104:7301–7306.
61. Bettencourt LMA, Lobo J, Strumsky D, West GB. Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities. PLoS ONE 2010; 5:e13541.

62. Jacobs J. The Death and Life of Great American Cities. New York, NY: Random House, 1961.
63. Sampson RJ, Morenoff JD, Gannon-Rowley T. Assessing "neighborhood effects": social processes and new directions in research. Annu Rev Sociol 2002; 28: 443–478.
64. Glaeser EL, Kallal HD, Scheinkman JA, Shleifer A. Growth in cities. J Pol Econ 1992; 100:1126–1152.
65. Madera C. The future of cities in the Internet era. Next City, 2010. Available online at http://nextcity.org/daily/entry/the-future-of-cities-in-the-internet-era
66. Harris JM, Hirst JL, Mossinghoff MJ. Combinatorics and Graph Theory. New York, NY: Springer, 2008.
67. Schläpfer M, et al. The scaling of human interactions with city size. Available online at http://arxiv.org/abs/1210.5215
68. Arrow KJ. A difficulty in the concept of social welfare. J Pol Econ 1950; 58:328–346.

Address correspondence to:

Luís M.A. Bettencourt
Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501

E-mail: bettencourt@santafe.edu

Appendix: The Planner's Problem

Definitions

Consider a city with N people defined as a connected social network.[56] This means that every person in the city can be connected to any other through a series of social links between third parties, however long. The social network that characterizes the city is G, which is made up of K social links. Both the general structure of G and its size K are functions of N and of other urban factors such as the costs (e.g., transportation, crime, time) and benefits (wages, information) of social interaction in that city.

The city's "performance" is monitored in terms of a set of quantities $Y_1, Y_2, \ldots, Y_M$. These urban metrics may vary from case to case, but I simply require that they are measurable quantitatively. The important property of each is that it is a function of the city's social configuration, that is, $Y_i(G)$ The metrics Y_i can measure familiar things, such as the size of the city's economy (GDP), the state of its public health, urban crime rates, the cost of transportation, and indices of environmental sustainability, including more subjective quantities such as happiness or satisfaction. I show below that what is measured and how many indicators are considered is not substantially important. The defining property is that they are functions of the social networks that make up the city, G.

Problem Statement

Given these definitions, the planner's problem is to choose the plan associated with the best G, subject on K remaining constant, so that a suitable function of the city's metrics Y_i is maximized.

Sketch of Proof

The planner's problem is computationally intractable: Without more specific knowledge, consider a city of N people characterized by all possibilities for social connections between them, K_P. For example, in the simplest case that these are undirected links, we have $K_P = N(N-1)/2$. Within this space the planner needs to determine the plan that assigns K connections optimally. The number of plans, P, therefore is given by the number of combinations of K in K_P (according to Harris et al.[66]) or

$$P(K_p, K) = \frac{K_p!}{K!(K_p - K)!} = e^{\ln K_p! - \ln K! - \ln (K_p - K)!} \equiv e^{B(K_p, K)} \tag{A1}$$

For large K_P, K, I can use Stirling's approximation to the factorial

$$\ln n! = n(\ln n - 1) + O(\ln n), \tag{A2}$$

in the exponent $B(K_P, K)$. After some algebra, I obtain

$$B(K_P, K) \simeq K\left[\ln\left(\frac{K_P}{K}\right) - 1\right] \simeq K \ln\left(\frac{K_P}{K}\right). \tag{A3}$$

Now, urban scaling theory[57] and cell phone data[67] tell us that $K = K_0 N^{1+\delta}$, where K_0 is a constant of order of a few connections, and $\delta \simeq 1/6$. The number of potential connections may vary somewhat depending on additional constraints on the problem, but the simplest assumption, given above, is that $K_P \sim N^2$ (Metcalfe's law).

Thus, with these choices, I obtain

$$B(N) \simeq B_0 N^{1+\delta} \ln N. \tag{A4}$$

Then, the number of plans that need to be evaluated to make the best choice is

$$P(N) \simeq P_0 e^{N^{1+\delta} \ln N}, \tag{A5}$$

which grows *faster than exponentially* with the number of people in the city. This is an astronomical number; for example, for a million people ($N = 10^6$), I obtain

$$P(N) = P_0 e^{1.4 \times 10^7} = P_0 10^{6 \times 10^6}, \tag{A6}$$

much, much greater than all the atoms in the universe ($\sim 10^{82}$).

The number of computational steps needed to evaluate the best plan is set by the evaluation of the quantities Y_i for each plan. Regardless of details, this can be done in $\sim M \times P$ steps, plus the sorting of these plans by these values. The fastest sorting algorithms require at least P steps ($P \log P$ is more typical), so I conclude that the complete algorithm would have a computational complexity proportional to at least $P(N)$. Such algorithms would scale with the population size of the city N, faster than exponentially, and would be impractical in all but the smallest towns.

Algorithms that scale exponentially or faster with the size of the input set are considered technically intractable; that is, they cannot be evaluated in practice as this set gets large. Thus, the (computational) planner's problem under the present set of assumptions is intractable.

Notes

1. The detailed combinatorial arguments involved in a specific algorithm for the planner's problem are likely to be more constrained if we include considerations of time, budget, and so on. This may result in effectively a smaller number of potential connections being viable, though it is uncertain whether such a judgment can be made *a priori*, without the consideration of all possibilities as done above. In any case, note that the leading term in the exponent B is set by the number of social connections K in the city, not K_P. As such, these considerations play a subleading role in the calculation.
2. Given a number of urban indicators, $M \geq 3$, there are well-known difficulties in producing an ordering of improvements that satisfies most inhabitants in the city. This is related to Arrow's impossibility theorem.[68] Such problems arise in addition to the computational complexity of the planner's problem.
3. Decisions and actions by individuals and organizations based on information they obtain in their urban environments effectively solve the planner's problem. They do so by "parallelizing" the problem and its inference and computations through the simultaneous pursuit of local adaptations that, in principle at least, can maximize each agent's preferences under constraints (budgets, knowledge, time, etc.). This self-organizing dynamics is not guaranteed to produce the best outcomes citywide, though. A more formal argument for self-organization in cities and what it can and cannot achieve remains an open problem that includes, but also needs to transcend, some of the foundational results in microeconomic theory.