WINNER *of the* 2013 EDWARD W. SOJA PRIZE

BIG DATA:

The Promise of a "Computational" Social Science

CHIRAG RABARI
University of California, Los Angeles

INTRODUCTION

In a 2009 paper in Science magazine, a cross-section of prominent academics and researchers surveyed the rapidly evolving landscape of digital technologies and forecast a revolution. Network systems and sensors integrated into the physical urban fabric, ubiquitous mobile connectivity, and omnipresent social networking platforms had given rise to unprecedented "Big Data" about the complex, interdependent components and mechanisms of the modern world. The authors stated their belief that a new field was emerging that could compile the traces of our digital transactions into "comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies" (Lazer et al. 2009, 721).

Such big data practices had already been occurring for several years at government agencies like the National Security Agency (NSA) and Internet companies such as Google and Facebook. Recent revelations about a vast global dragnet mining the social connections of U.S. citizens indicate these activities have only accelerated over time (Risen and Poitras 2013). But the researchers hoped these data sets and the skills to interpret them would not remain solely in the realm of governments and private corporations. To serve the public good of advancing knowledge, they laid out an agenda for what they termed "Computational Social Science" (CSS).

In this seminal piece and in subsequent complementary articles by different contributors, the authors made the case that CSS will offer new perspectives on individual and collective human behavior, with attendant possibilities to increase human welfare through the enhanced prediction and control of our social systems. In particular, a big-data-driven CSS will enable the design of systems that are "more stable, fair, and efficient" than our current, classically influenced paradigms of human society (Pentland 2012). Some of the researchers claimed that big data is a new scientific tool on par with the introduction of the microscope in the sixteenth and seventeenth centuries, bringing previously obscure social processes into clear sight.

The city is emerging as both subject and object of this revolution. Big data and CSS are predicted to increasingly inform the future management and governance of cities as well as the interactions and experiences of people who live in them; the city is also the source of and platform for the data that will generate these novel uses (Rabari and Storper 2013).

Urban planning, as an applied, cross-disciplinary social science field, is likely to be particularly influenced by these new practices and paradigms. The discourse is becoming crowded with hopes and predictions that we will be able to make our cities more efficient and livable, creating a rush to adopt and implement new technologies. New specialties informed by CSS, such as Urban Science and Informatics, are already growing quickly.[1]

None of this is surprising—major technological innovations often lead to the capture of attention and resources. Such developments, however, call for the thoughtful engagement of urbanists and critical theorists. The emergence of new fields of inquiry is no small matter, and the development of new methodological paradigms should be treated with rigor and care. There are many implications for social science, in particular the problems that will be identified and addressed, and the questions that will be asked and answered. But these implications are as yet scarcely analyzed or understood.

The computational social scientists have thus far staked out a number of explicit and implicit claims about the nature of knowledge, the proper conduct and purpose of science, and human beings and their various forms of collective organization. Many of these arguments can be situated within long-standing debates in the philosophy of social sciences, including questions of how best to study society, what methodologies

[1] The most prominent of these efforts is the new Center for Urban Science + Progress at New York University, which opened in the fall of 2013.

are most appropriate for scientific inquiry, and the problems of social control.

This paper seeks to untangle some of these foundational issues by offering a critical analysis of CSS as a nascent field of inquiry within the wider social sciences literature. It is organized in the following manner:

- In Section 1 we will examine the anticipated benefits of this emerging field, and obstacles to its potential success.

- In Section 2 we will unpack the epistemological issues within the authors' claims, paying particular attention to the positive, normative, and methodological components of the CSS agenda.

- In Section 3 we will assess the viability of CSS as a scientific research programme based on the criteria of the noted mathematician and philosopher of science Imre Lakatos.

We will close with some thoughts on big data and CSS through the narrative of social-scientific "progress," and assess where these narratives intersect with issues of particular concern to urban planners and policymakers. The key takeaway from this investigation is that, although CSS will enable impressive and important gains, these gains will necessarily be limited to particular types and areas of inquiry. Where CSS does in fact suggest "revolutionary" possibilities, it will raise issues that are fundamentally political and moral in character, and therefore outside the domain of technologically driven scientific inquiry.

1. BENEFITS *and* OBSTACLES *to a* COMPUTATIONAL SOCIAL SCIENCE

Lazer et al. (2009) specify three primary benefits to a CSS, each with its own ancillary implications. (These core arguments have been expounded upon on by many of the individual scholars as well; see, for example, Barabasi 2012; Christakis 2012; Pentland 2012.)

> (i) "To date, research on human interactions has relied mainly on one-time, self-reported data on relationships. New technologies, such as video surveillance, email, and 'smart' name badges offer a moment-by-moment picture of interactions over extended periods of time, providing information about both the structure and content of relationships" (Lazer et al. 2009, 722).

A CSS could thus offer greater insight into the temporal dynamics of our communication, behavioral, and proximity patterns, leading to predictions about individuals or collectives in specific relational or interaction contexts.

> (ii) "We can also learn what a 'macro'

> social network of society looks like, and how it evolves over time" (Lazer et al. 2009, 722).

This macro view—either through the records of phone companies, the data of search and commerce sites like Google or Amazon, or the large-scale tracking of people's movements or transactions—could offer a comprehensive view of societal-level patterns of communication, transportation, economic activity, or health and epidemiology.

> (iii) "The Internet offers an entirely different channel for understanding what people are saying, and how they are connecting… Virtual worlds… by their nature capture a complete record of individual behavior" (Lazer et al. 2009, 722).

The power of this kind of data is that it offers information about people's behaviors instead of their beliefs, on the assumption that what a person actually does is more important than what they think (Pentland 2012). Indeed, it is believed that individuals could have "much of their life, almost in minute resolution…reconstructed from the many data streams [they] leave around [them]" (Barabasi 2012).

In summary, the big data behind a computational social science involves not just aggregating all social patterns, but simultaneously disaggregating at the level of agents. This data would encompass millions of people and entities and be high in spatial, temporal, and typological resolution. This would allow scientists first to describe, then to quantify, then to formulate, then to predict, and finally—possibly—to even control complex systems such as the economy or society (Barabasi 2012).

There are, however, substantial institutional barriers that could limit the development of a CSS, which are characterized and summarized as follows:

(i) The possible inadequacy of current methodological and theoretical paradigms to study and analyze data of this breadth, depth, and scale.

(ii) The challenges to observation inherent in studying human subjects as compared to the natural world.

(iii) Underdeveloped infrastructure within the social sciences.

(iv) Concerns about access, privacy, and ownership.

(v) The need to train new, interdisciplinary scholars who are comfortable working across multiple fields (Lazer et al. 2009).

To illustrate, Lazer et al. (2009) note how existing sociological network theory was developed with "snap-shot" data built upon relatively limited samples; we now have longitudinal data sets that could conceivably include every single location, financial transaction, and communication of millions of people. There may be significant limitations to what traditional theories can tell us and to the skills within social science departments to make sense of this data, as well as barriers of ownership and access to the relevant data sources, many of which are proprietary and balkanized.

Although such concerns are important, they have not been meaningfully explored. There are also a number of other important questions that CSS proponents do not engage with, as well as a significant body of knowledge that have not been addressed.

2. BIG DATA, COMPUTATIONAL SOCIAL SCIENCE, *and the* PHILOSOPHY *of* SOCIAL SCIENCE

What then are the primary epistemological issues at play here? Although it is difficult to precisely situate these computational social scientists within any particular school or tradition, they do, however, have certain key affinities that can be identified.

First, EMPIRICISM is privileged as the primary means of understanding the world and computational DATA is treated as the most accurate or best possible approximation of reality. In this case, social reality is simply what is recorded and quantified by our technology.[2] We are now capturing more complete data than ever before, and this data is by definition useful in terms of the insights that can be gleaned from it.

This is what the authors Viktor Mayer-Schonberger and Kenneth Cukier refer to in their book Big Data: A Revolution That Will Transform How We Live, Work, and Think as the move from limited samples to something approaching "N=all" (2013).

Further, these huge data sets are reliably meaningful because we presuppose the existence of INTELLIGIBLE, RATIONAL BEHAVIOR which can be divined at the individual, group, organizational, or societal level (D'Agostino 2011).

Second, advanced STATISTICAL and PROBABALISTIC methods are heavily leaned upon as the primary analytic and methodological tools at the social scientists' disposal. This yields a major focus on the discernment of patterns, or CORRELATIONS, which are deemed sufficient for action, explanation, or decision-making.

[2] For example: "What those [digital] breadcrumbs tell is the story of your life… Who you actually are is determined by where you spend time, and which things you buy" (Pentland 2012), or "This is the first time that we can know what people are doing in an objective manner, without biases, without lying, without kidding ourselves, of trying to present a different image than what we are" (Barabasi 2012).

The fact that the data might be "messy," or that we may not be able to precisely determine causation, is not of major concern; the "what" will be more important than the "why" (Mayer-Schonberger and Cukier 2013).

Further, it is believed that if theoretical insights and frameworks are to emerge, they will arrive *inductively* from the richness of these new data sets (Barabasi 2012; Pentland 2012).

And finally, *prediction* and other *instrumentalist*[3] goals are treated as the primary aim, or ends, of the scientific process. There is thus a great desire for the production of "technically utilizable" knowledge (Habermas 1976) that can serve as "calculating devices" (Lakatos 1999), often for the purposes of enabling specific interventions and exerting *control* over the environment.[4]

There are many examples where big data has already been put to this kind of use in urban social science contexts: models that can help determine when bridges or other infrastructure are under dangerous levels of stress, when abandoned properties might represent fire hazards, how to best reroute city bus lines to yield efficiency gains, or where flu outbreaks might be occurring and how best to intervene, and so on (Mayer-Schonberger and Cukier 2013).

So: how can we better contextualize these affinities within the wider social sciences literature?

2.1

The first and most important question concerns how to study society. The basic premise of the computational social scientists is that we now have more data than ever before; this data represents something, and our task is to determine its instrumental usefulness. The social sciences, however, have historically been concerned not just with "data" but also with matters of meaning, interpretation, understanding, and reflexivity. This has been the subject of a long-standing debate between empiricists, who sought to move beyond the problems of subjectivity, and hermeneutists, who placed subjective experience at the core of social reality (Taylor 2001). In hewing so

[3] Here we mean means-ends or problem-solution oriented thinking

[4] For example: "Do you want to stop different transmitted diseases? Do you want to design better cities? Do you want to stop traffic jams? The data to do so is there in private hands, and we need to identify some social consensus by which the data can be shared with the different stakeholders who can take advantage of that," (Barabasi 2012), or "The fact that we can now begin to actually look at the dynamics of social interactions and how they play out, and are not just limited to reasoning about averages like market indices is for me simply astonishing. To be able to see the details of variations in the market and the beginnings of political revolutions, to predict them, and even control them, is definitely a case of Promethean fire. We're going to reinvent what it means to have a human society" (Pentland 2012).

closely to the empiricist tradition with its emphasis on "brute data," the practitioners of CSS open themselves to a number of well-worn critiques.

First, despite the fact that scholars may claim to "let the data speak for itself," such an act is impossible. Data does not "speak"—construction, organization, and interpretation are all crucial components of quantitative modes of inquiry (Clarke and Primo 2012). We must choose what to measure and how to measure, and particularly in the case of new social technologies we are dealing with many self-constructed categories such as "likes," "friends," "shares," "tweets," and so on. How can we know for sure that we are choosing the correct empirical frame of reference in our investigations? How shall we address the fact that our data will be extremely rich for certain categories of activity, which will necessarily become the objects of our study, and nonexistent for others that we have not formally conceptualized or systematized?

In the past, critical theorists argued that exclusively empirical methodologies have a tendency to legitimate "facts" about society while circumscribing the language to talk about intractable or obscure social phenomena, such as justice, power, or exploitation (Adorno 1976), or the entire category of intersubjective and common meanings. This "brute data" may be fine in a natural sciences context, and indeed the philosopher Charles Taylor noted how the "the great achievement of the seventeenth-century scientific revolution was to develop a language for nature that was purged of human meanings" (Taylor 2002). This explanatory language represented an advance of knowledge because it enabled some finality in terms of understanding the objects of our inquiry. Such a stance is problematic in the social sciences, however, because we can never assume a "final" understanding when engaged in communicative discourse with an interlocutor. Concepts, creativity, rules, conventions, beliefs—these are all essential to human behavior. Our understanding of meaning as applied to action is therefore partially constitutive of the reality that we study, as are the potential, irrational gaps between our self-understanding of our motives and values and our actual behavior, and the gaps between what is and what could be within the social order (Fay and Moon 2001). A big data methodology would not, for example, be able to say anything about an idea or innovation that hasn't been thought of yet.

On this front the proponents of CSS seem very concerned with what people do—what is indicated by and recorded of their surface-level interactions—and not at all concerned with the interplay between computational data, a person's interior life, and the wider, intricately complicated, constantly evolving social reality that informs both.

This raises another issue that is not explicitly addressed in the CSS agenda: the age-old question of whether individuals are shaped by societies or societies are merely the aggregate results of individualistic, "bottom-up" processes. A CSS devoted to developing comprehensive pictures of individual and group behavior would necessarily require a perspective on the precise nature of our "social" minds; it would otherwise be impossible to establish a framework or starting scale to make sense of whatever correlative patterns emerge from the data we investigate.

Historically the social sciences have seen five general (not exhaustive) approaches that have tried to answer this question, each with their own advantages and limitations: individualist, social, extended, collective, and the "social brain"[5] (Kaufmann 2011). Based on their writings so far, the assumptions of CSS seem to sit somewhere within the individualist and collective approaches, where in the former there is no a priori "community" to assess; instead collective phenomena are just the aggregate results of rational actors intelligibly pursuing their wants and preferences.[6] In the latter case, Kaufmann (2011) identifies an instrumentalist approach that he terms "non-distributive collective attribution." Here we say that collective "groups" exist not in any metaphysical or causal sense, but that such attributions "fulfill explanatory, interpretative, and predictive needs….This intentional stance does not necessarily involve any ontological commitment: it is above all a good tool for making a priori preferences and behaviors intelligible and predictable" (Kaufmann 2011, 169).

Although these approaches have their relevance and justifications, they each entail their own positive, normative, and methodological inclinations. They thus leave out

[5] Kaufmann defines these as follows. Individualist approaches to the social "see the social as a matter of mental contents, either conscious or unconscious, which result from the translation of the social world around the mind into representations." Social approaches to the mind emphasize "the social dimension of any perceptions, expectations, categorizations, emotions, and representations that furnish the human mind. Subjective minds are nothing but the holders of common meanings and impersonal rules that constitute the objective mind of a given community." The extended mind approach "calls into question the demarcation of skin and skull, and the assumption that what is outside the body is also outside the mind". Collective approaches "also advocate that minds are not 'in the head,'" but rather operate through groups. Approaches to the "social brain" draw from "developmental, comparative, and evolutionary psychology the hypothesis according to which social species are endowed with pre-wired, well-adapted cognitive devices to process specific ontological domains" (2011).

[6] According to MIT's Alex Pentland (2012), for example, Adam Smith and Karl Marx were wrong. Why? "Because they talked about markets and classes, but those are aggregates. They're averages….While it may be useful to reason about averages, social phenomena are really made up of millions of small transactions between individuals… You need to get down into these new patterns, these micro-patterns, because they don't just average out to the classical way of understanding society. We're entering a new era of social physics, where it's the details of all the particles—the you and me—that actually determine the outcome".

a lot of the richness of varying different accounts of individual and social behavior. A comprehensive effort would need to do more to incorporate and adjudicate between the many different approaches to "socialness," with an eye towards establishing how different computing technologies enable, express, or condition social behavior.

2.2

A second major issue surrounds questions of method and how to "do" science. On this question Lazer et al. (2009) were fairly open about what they see as their primary methodological and scientific challenges, almost all of which surround coping with the "data deluge." For example, Alex Pentland (2012) of MIT claims that the traditional scientific method is no longer usable because there will be so much data that it will be very difficult to unpack causality. Albert-Laszlo Barabasi (2012) of Northeastern University pushes along similar lines, noting that science is becoming a "byproduct" of all the data we have at our disposal, and that we need a "new science" to make sense of these challenges.

These are striking claims. If adherents of CSS are to move beyond simply identifying correlations for predictive purposes, these comments suggest something of an "inductive turn" away from the hypothetico-deductive (H-D) method upon which much of the modern scientific revolution was built.

> Instead of the incremental process of falsification of previous hypotheses and deductive, structured approaches to information, we can now take advantage of abundant computing power to "see what the data says" in any direction. We thus could be on the verge of discovering relationships and patterns which were obscured by deductive epistemology, allowing us to reach a new level of understanding human social life. (Rabari and Storper 2013, 17)

The hope of being able to generalize theories after the accumulation of sufficient data has been a dream of science for a long time, and in fact brings us all the way back to Francis Bacon and the seventeenth century (Hollis, 1994). But it is important to note that over the history of science, induction never succeeded as an operative scientific paradigm, despite some noted adherents such as Newton (that Newton did not really arrive at his laws through induction has been well established, see, for example, Lakatos 1999). Inductive epistemology faced many hurdles that caused it to be abandoned as a "degenerative" research programme in favor of the now-standard H-D approach: infinitely many valid inferences, infinitely many axioms, its tendencies towards content preservation, the inability to adjudicate between varying explanations, and so on (Lakatos 1999).

The computational social scientists have acknowledged some of these underlying challenges, but they have not yet articulated what the solutions might be. Do we need entirely new theoretical frameworks or can our existing ones be effectively repurposed? How will big data analysis actually help us resolve micro and macro levels of investigation and move systematically from agent-based activities to emergent macro-level social phenomena? Beyond the implicit exception of rational intelligibility as an assumed feature of technologically mediated interactions, there has thus far been much more discussion around the building and use of empirical models for the purposes of prediction, measurement, and characterization as opposed to theoretical models which can fulfill foundational, organizational, or exploratory roles (Clarke and Primo 2012).

Answers may arrive through technical innovation and new, yet-to-be-invented methodologies; or, as previously noted, scholars may instead choose to focus solely on probabilistic or instrumentalist goals. In this scenario scientists would not concern themselves with the truth or falsity of particular theories, or with the systematic judgment of how one theoretical framework is better than another, or with the entire matter of causation in social affairs, but mostly with calculation, prediction, and the unwinding of significant correlations. This type of project could address one of the social sciences' greatest historical limitations (their inability to make verifiable predictions) at the expense of several other significant problematics.

2.3

A final major issue concerns the problems of predictive control. Some commentators remarked on the privacy issues raised by big data, or the "Promethean fire" that new legibility of social processes could represent, but these issues were not systematically addressed. Given the instrumental bent of the current CSS project, it seems likely that researchers (or governments, or corporations) could eventually achieve success within specific realms and contexts at predicting and thereby controlling complex human systems and individual human beings. These are very powerful tools that are being developed, but there is reason for caution in assuming they will automatically be used for the betterment of human society. The aforementioned NSA spying scandal is a perfect example of the dangers of big data methods and technology divorced from critical discourses about freedom, justice, and power.

What this illustrates is that the kinds of mechanisms at work in our society depend partly on what we decide to set up, or allow; "the question of the implementation of social science is thus partly the issue of how much control we want to impose of human behavior" (Guala 2011, 590). This issue—the centrality of politics and morality as reflexive components of the world that the social sciences seek to understand—is again unaddressed within the CSS agenda.

Taylor, in his "Interpretation and the Sciences of Man," gives perhaps the best treatment of the problems of prediction in a social science context that gives pride-of-place to meaning, interpretation, and subjective experience. First is the "open system" predicament, whereby human events, like meteorology or ecosystems, cannot be protected from external interference; second, Taylor points out that a science of interpretation can never achieve the exactitude of measurement seen in the natural sciences; and finally, he notes that man is a "self-defining" animal, prone to incommensurable "conceptual mutations" (Taylor 2001).

For our purposes, these objections help us see that big data could help us react to and "fix" imminent problems, but these actions would necessarily have to be targeted and narrow. By definition it will not be able to tell us about the wisdom or effects of significant societal interventions. This is because the "n-space" of variables is always changing, particularly after we act. Those actions in turn change what is measureable and possible.

> The success of prediction in the natural sciences is bound up with the fact that all states of the system, past and future, can be described in the same range of concepts, as values, say, of the same variables….This conceptual unity is vitiated in the sciences of man by the fact of conceptual innovation, which in turn alters human reality….Really to be able to predict the future would be to have explicated so clearly the human condition that one would already have preempted all cultural innovation and transformation. (Taylor 2001, 209).

These issues have a well-known analogue in computer science, sometimes referred to as a "decision problem." In pointing out the fatal flaws in the ideal of a security state, science historian George Dyson (2013) discussed how there is no way to determine what any particular line of code in a complex system will do without letting it run. Thus, "no firewall that admits anything new can ever keep anything dangerous out… Any formal system that is granted (or assumes) the absolute power to protect itself against dangerous ideas will of necessity also be defensive against original and creative thoughts".

The efficacy of big-data-driven predictive control over large-scale, complex systems could thus depend on a "frozen" world where all variables and contingencies are known beforehand. The problems this presents for human flourishing and freedom are not difficult to imagine.

3. THE VIABILITY *of* CSS *as a* SCIENTIFIC RESEARCH PROGRAMME

Proponents of CSS see themselves as defining a new, interdisciplinary field. Their model is the emergence of cognitive science that bridged subjects such as philosophy, computer science, and neurobiology. Does this field, as it has been defined thus far, offer the promise of meaningful scientific progress?

The work of the noted mathematician and philosopher of science Imre Lakatos can offer some insights here. Lakatos was above all concerned with the issue of demarcation: how to distinguish good science versus bad science, theoretically progressive versus degenerating problem shifts, good verifications versus bad ones, good dogmatism versus bad dogmatism, and novel facts from useless ones. Lakatos built upon the works of Karl Popper and Thomas Kuhn to create a methodology of scientific research programmes to help us separate "the goodies from the baddies" (Lakatos 1999).

A research programme, in his view, "sets out the fundamental conceptual framework or conceptualization of the phenomena we wish to explain, and the rules in accordance to which theoretical innovations or developments will be made" (Fay and Moon 2001, 30). This methodology could help explain why the Newtonian paradigm of the universe, despite its many anomalies and eventual supersession by Einstein, was still theoretically progressive, "good" science, because it led to unexpected findings and therefore better questions. Astrology and Freudianism, however, were hopelessly pseudo-scientific.

The social sciences have always presented several problems for this view, problems that may or may not be mitigated by the increasing quantification of social life.

> The question immediately arises whether the intentional nature of social phenomena constrains what can count as an adequate research program… [I]t suggest that an adequate explanation of a social phenomenon would have to include, or be based upon, an account of the reasons or motivations which led to the behavior which brought about the phenomenon in question…. Research programs in the social sciences would have to include a conception of human needs, purposes, rationality, etc. in terms of which these motivational accounts could be constructed. (Fay and Moon 2001, 30)

On this reading the prognosis for CSS is unclear. Beyond a commitment to empirical data, limited rationalist conceptions of individual and collective intentions, a focus on specific analytic and methodological tools, and instrumentalist ends, it is unclear what makes up the actual core of this research agenda—what its basic theoretical assumptions are about social life, and what it is trying to discover. These insights may emerge inductively from the massive amounts of data now available, but this will run counter to the modern history of scientific progress, where intuitions about where to look and what to look for led to a progressively expanding penumbra of knowledge. In fact, traditional theoretical models may prove more important than ever in the new environment, as the signal-to-noise ratio increases exponentially.

That being said, the combinations of raw empiricism, rationalism, and instrumentalism matched with powerful technology could be enough. With a bit of tenacity and dogmatism, CSS may indeed prove to be a progressive paradigm. The catch, however, is that there is no guarantee this progressivity would educate us about meaningful social imperatives. For example, an in-depth big-data investigation of a social networking platform like Facebook may illustrate more about Facebook than it does about the wants, needs, motivations, and values of the people using it.

The danger exists that this would be a reified progressivity; that these tools and systems of inquiry—which embody their own logics and intentionalities within the routinized structures of our socio-technological systems—could outrun or overrun the original functions they were designed to serve and support (Wolin 1969). In short, computational social science could say more about technology than it can say about us.

4. CONCLUSION

What we hope to have illustrated thus far is that big data and CSS present something of a mixed bag when we consider the topic of scientific progress. The absence of a theoretical core, the problems of induction, the limitations of purely data-driven approaches, the staggering complexity of social behavior and social systems, and the moral and political questions raised by the problems of predictive control should all serve to temper some of the current, mildly utopian hopes about what will be possible.

All of that said, big data and CSS still represent a clear advance when viewed as a technical phenomenon; there really is a vast process of “datafication” occurring in the world right now (Mayer-Schnoberger and Cukier 2013). On the social-scientific front we can expect the creation of many interesting and sophisticated empirical models that will allow us to identify significant correlations in the data we’ve gathered. Recognizing these patterns will in turn allow us to intervene in specific ways to help us solve specific problems, whether that means improving efficiency, identifying pandemics, or better managing our infrastructure.

However, while efficiency, improved management, and the discernment of patterns are important goals, for many urbanists and critical theorists they will not be enough, especially if they crowd out others areas of focus and discussion. Technological revolutions, like any new, shiny thing, can sometimes distract us from the difficult, hard problems, such as poverty, development, or social justice, which were the reasons many of us were called to the social sciences in the first place.

To illustrate by way of example: using a big-data-driven CSS model to create a 10% efficiency improvement on a bus line is a real, serious utility gain. These kinds of small, meaningful improvements are the "bread and butter" of urban planning, where we, as practitioners, can make a difference in people's everyday lives. But it is still a limited conception of public service and civic-mindedness, especially if these efficiency gains are happening while fares are being raised and services being cut. In these instances there are perhaps other conversations we should be having and other problems we should be focusing on.

Of course, as planners we will want to refine our skills and make sure that we are literate in all of the new tools and methodologies as CSS practices infiltrate the field. But there is a useful analogy here. While Geographic Information Systems (GIS) have been standardized and fully incorporated into social science departments across the country, and has had major effects on our "praxis" as planners, it can hardly be said to have created a revolutionary impact on the wider socio-economic system (Rabari and Storper 2013).

This relates to a very important point: the wider social, political, and economic context in which these new technologies will be applied matters a great deal, but proponents of data-driven social science almost never discuss these serious foundational problems. The United States has seen stagnating incomes, massively increased inequality, huge sorting along demographic and regional lines, and inexorable political gridlock. Big data could allow us to "see" these phenomena better, but it could also just codify them or even exacerbate them, depending on the kinds of decisions being made and by whom. For example, we may see new issues where patterns of social privilege are embedded but invisible within data sets themselves, perhaps due to the differing propensities of different social groups to participate in "datized" moments, or because of unequal abilities to interpret and make effective use of data (Johnson 2013). The fact is that we simply don't know, and a purely instrumentalist project will by definition not help us understand, especially one that limits the objects of social scientific inquiry to outward, quantifiable expressions of social life.

There is a reason that the majority of real-world big data applications seen thus far have been in the realm of state spying and advertising (Mayer-Schonberger and Cukier 2013). These are easy, low-hanging fruit, and they fit neatly into existing power structures, whether bureaucratic or financial. The access, knowledge, and means to exercise these sorts of sophisticated projects of technical control will always be closely guarded by those who have a lot to gain or a lot to lose.

All of which is to say: the struggle for justice, the never-ending drive for reform, and the vital role of cultural, legal, and social norms will have to remain front and center to any critical urban theory that seeks to deal with the world as it is, and as it might be. The dramatic failures of 50s and 60s era rational planning should loom large in our minds. Technocratic approaches have their limits.

Many thanks to Karna Wong, Brian Esparza Walker, and Michael Storper for their direction, assistance, and feedback.

Chirag Rabari is a recent graduate of the MURP program at UCLA. His current research focuses on the impact of emerging digital technologies on management, governance, and social interaction within cities. Originally from Massachusetts, he has past professional experience in the fields of journalism, financial technology, and open-source project management.

Adorno, T. 1976. Sociology and empirical research. In The Positivist Dispute in German Sociology, ed. G. Adey and D. Frisby, 68-86. London: Heinemann.

Barabasi, A-L. 2012. Thinking in network terms. Edge, September 24. http://edge.org/conversation/thinking-in-network-terms.

Christakis, N. 2012. A new kind of social science for the 21st century. Edge, August 21. http://edge.org/conversation/a-21st-century-change-to-social-science.

Clarke, K., and Primo D. 2012. A model discipline: Political science and the logic of representation. New York: Oxford University Press.

D'Agostino, F. 2011. Rational Agency. In The SAGE Handbook of the Philosophy of Social Sciences, ed. I. Jarvie and J. Zamora-Bonilla, 182-198. London: SAGE Publications.

Dyson, George. 2013. "The Entscheidungsproblem." Frankfurter Allgemeine Zeitung, 7, sec. Feuilleton. http://www.faz.net/aktuell/feuilleton/debatten/ueberwachung/the-military-industrial-complex-the-entscheidungsproblem-12306230.html.

Fay B., and Moon, D. 2001. What would an adequate philosophy of social science look like? In Readings in the Philosophy of Social Science, ed. M. Martin and L. McIntyre, 21–35. Cambridge, MA: MIT Press.

Guala, F. 2011. Experiments. In The SAGE handbook of the philosophy of social sciences, ed. I. Jarvie and J. Zamora-Bonilla, 577–593. London: SAGE Publications.

Habermas, J. 1976. The analytical theory of science and dialectics. In The Positivist Dispute in German Sociology, ed. G. Adey and D. Frisby, 131-162. London: Heinemann.

Hollis, M. 1994. The philosophy of social science. Cambridge, UK: Cambridge University Press.

Johnson, J. A. 2013. From open data to information justice. Midwest Political Science Association Annual Conference, April 2013 Version 1.1. http://

ssrn.com/abstract=2241092.

Kaufmann, L. 2011. Social minds. In The SAGE handbook of the philosophy of social sciences, ed. I. Jarvie and J. Zamora-Bonilla, 153–181. London: SAGE Publications.

Lakatos, I., and P. Feyerabend. 1999. For and against method: Including Lakatos's lectures on scientific method and the Lakatos-Feyerabend correspondence. Chicago: University of Chicago Press.

Lazer, D., A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, and M. Gutmann. 2009. Computational social science. Science 323: 721–723.

Mayer-Schonberger, V., and K. Cukier. 2013. Big data: A revolution that will transform how we live, work, and think. New York: Houghton Mifflin Harcourt.

Pentland, A. 2012. Reinventing society in the wake of big data. Edge, August 30. http://www.edge.org/conversation/reinventing-society-in-the-wake-of-big-data.

Rabari, C., and M. Storper. The digital skin of cities: Urban theory and research in the age of the sensored city, ubiquitous computing, and big data. Cambridge Journal of Regions, Economy and Society (forthcoming).

Risen, J., and L. Poitras. 2013. N.S.A. gathers data on social connections of U.S. citizens. The New York Times, September 28. http://www.nytimes.com/2013/09/29/us/nsa-examines-social-networks-of-us-citizens.html.

Taylor C. 2002. Gadamer on the human sciences. In The Cambridge Companion to Gadamer, ed. R. Dostal, 126–142. Cambridge, UK: Cambridge University Press.

Taylor C. 2001. Interpretation and the sciences of man. In Readings in the Philosophy of Social Science, ed. M. Martin and L. McIntyre, 181–211. Cambridge, MA: MIT Press.

Wolin S. 1969. Political Theory as a Vocation. The American Political Science Review, 63 (4), 1062–1082.