

Locative media and data-driven computing experiments

Big Data & Society
 January–June 2016: 1–12
 © The Author(s) 2016
 Reprints and permissions:
sagepub.com/journalsPermissions.nav
 DOI: 10.1177/2053951716652161
bds.sagepub.com



Sung-Yueh Perng¹, Rob Kitchin¹ and Leighton Evans²

Abstract

Over the past two decades urban social life has undergone a rapid and pervasive geocoding, becoming mediated, augmented and anticipated by location-sensitive technologies and services that generate and utilise big, personal, locative data. The production of these data has prompted the development of exploratory data-driven computing experiments that seek to find ways to extract value and insight from them. These projects often start from the data, rather than from a question or theory, and try to imagine and identify their potential utility. In this paper, we explore the desires and mechanics of data-driven computing experiments. We demonstrate how both locative media data and computing experiments are ‘staged’ to create new values and computing techniques, which in turn are used to try and derive possible futures that are ridden with unintended consequences. We argue that using computing experiments to imagine potential urban futures produces effects that often have little to do with creating new urban practices. Instead, these experiments promote Big Data science and the prospect that data produced for one purpose can be recast for another and act as alternative mechanisms of envisioning urban futures.

Keywords

Data analytics, computing experiments, locative media, location-based social network, staging, urban future, critical data studies

Introduction

Urban social life has rapidly and pervasively become geocoded, mediated, augmented and anticipated by location-sensitive technologies and services (Graham and Zook, 2013; Kinsley, 2011). Locative media in particular, ranging from the early platforms such as Brightkite and Gowalla¹ to more recent variations, including Foursquare, Facebook, Snapchat and Moves, have quickly been incorporated into everyday urban life as popular apps on smart phones and have increasingly attracted users to only a handful of networks. As of early 2014 Twitter has 243m users,² Facebook 131m active users and 68m mobile users (plus 900m less active users)³ and Foursquare 45m.⁴ Locative media provide users the functionalities of: sharing their everyday life and local knowledge in relation to places they visit, e.g. ‘checking-in’ to stores; leaving their comments about these places; sharing and geo-tagging photos, messages, updates and friends when going out together; or tracking daily mobility

patterns. They have become critical to how these people sustain and create social connections, how they filter and encourage the circulation of certain information, and how they reshape places they visit through their virtual and embodied visits (Evans, 2015).

By interacting with locative media, users generate a huge amount of granular data and metadata about themselves, their interactions and shared interests, and places they visit. Services such as Snapchat can, with consent, track the location data of photos and the devices used to share them, alongside a wide range of data about the usage (‘time, date, sender, recipient of a message, the number of messages you

¹Maynooth University, Ireland

²University of Brighton, UK

Corresponding author:

Sung-Yueh Perng, National Institute for Regional and Spatial Analysis, Iontas Building, Maynooth University, Maynooth, Co., Kildare, Ireland.

Email: sung-yueh.perng@nuim.ie



exchange with your friends'), the contents (photos and messages) shared, device details ('hardware model, operating system and version, unique device identifiers (including MAC address and IMEI), browser type and language, mobile device phone number, and mobile network information'), photos, contact lists, web browser histories and information from other tracking technologies.⁵ The use of such media then creates a massive amount of continuously updated 'locative' data about everyday personal, social, temporal and spatial practices.

This avalanche of social and spatial data has generated much excitement, expectation and imagination since they can be leveraged to design spatiotemporal models to explain, predict, simulate, and optimise spatial behaviour, transportation and economic activity (McArdle et al., 2014). However, concerns have been raised over the wider implications of such 'data-driven urbanism' – that is, new forms of real-time management and regulation of city services and infrastructures based on the processing and analysis of a deluge of urban Big Data (Kitchin, 2015). Locative media data can be used to identify personal characteristics, track and trace location and movement, and monitor activities, with these data used to profile customers and target advertising and marketing, enabling enterprises to understand consumption practices and build relationships with consumers (Evans, 2013). More broadly, many issues have been raised concerning the social, political, epistemological and ontological challenges, transformations and consequences associated with the generation, development and application of Big Data and data analytics in diverse domains (Barreneche and Wilken, 2015; boyd and Crawford, 2012; Leonelli, 2014).

These issues have led to calls for a critical understanding of the 'knowledge' derived from locative media data, the tools, techniques and technologies involved in such knowledge production processes (Wilson, 2015), and the ethics of data practices with respect to privacy and data protection and data security (Edwards, 2015; Kitchin, 2016).

Termed critical data studies, Dalton and Thatcher (2014) thus call for an extended programme of research into regimes of data and set out seven provocations they felt were needed to provide a comprehensive critique of the production, commodification, analysis and usage of data:

1. Situate data regimes in time and space.
2. Expose data as inherently political and whose interests they serve.
3. Unpack the complex, non-deterministic relationship between data and society.
4. Illustrate the ways in which data is never raw.

5. Expose the fallacies that data can speak for themselves and that Big Data will replace small data.
6. Explore how new data regimes can be used in socially progressive ways.
7. Examine how academia engages with new data regimes and the opportunities of such engagement.

Kitchin and Lauriault (2014: 6) contend that one way to operationalise such research is to:

unpack the complex assemblages that produce, circulate, share/sell and utilise data in diverse ways; to chart the work they do and their consequences for how the world is known, governed and lived-in; and to survey the wider landscape of data assemblages and how they interact to form intersecting data products, services and markets and shape policy and regulation.

Furthermore, Barnes and Wilson (2014) argue for the need to trace the history of Big Data and the assumptions and contestations that underpin their development.

While we believe realising these provocations and the suggested approach is productive, they do not circumscribe critical data studies and there are many other issues that need to be examined and other potential avenues of enquiry. One of these is to study how data are reimagined and repurposed and to explore the unanticipated uses of data and the unintended consequences that arise from such usage. This is the approach we take in this paper, examining how locative media data have been repurposed in diverse ways within computing experiments, often with the hope of creating novel ways to envision and pursue 'data-driven urbanism'. As such, the paper contributes to critical data studies by examining how locative data are experimented with to propose new services for and produce new knowledge about cities. In so doing, the paper presents a novel way of identifying, exploring, tracing and tracking the desires, mechanics and innovations of data-driven computing experiments, demonstrating how they seek to render locative data useful and actionable and repurpose them to create new value.

In particular, while recognising that data derivatives possess a unique ontology derived from the association within data segments (Amoore, 2011), we examine the 'unintended expectations' of computing experiments as a way of tracing the mutual configuration of data and techniques enacted and required when assembling them. Tarde (2007) and his conceptualisation of economy and passion are particularly useful here. For Tarde, economics is a science of both quantification and passion. To understand markets he argues that one has to recognise that both are 'made up of passions whose astonishing development... amplified their

interconnections' (Latour and Lépinay, 2009: 24–25). We can follow Tarde and argue that 'there is not a single aspect of social life in which one does not see passion grow and unfold together with intelligence', even including in the domain of science where 'passion and reason, from age to age, progress hand in hand' (Tarde, 2007: 631). As we will detail, computing experiments are similarly made up of quantification, passions and unintended expectations as data produced for one purpose is transformed for indeterminate causes.

By treating computing experiments as case studies, our aim is to capture and conceptualise the moments when the coupling of locative data and computing techniques take place and to ask what is enacted throughout such processes. Such a 'reading' of computing experiments is more than discourse analysis. Treating computing experiments as objects, we draw upon the anthropological tradition of recognising the agency of objects to further examine the critical steps through which locative data and computing experiments are made fit for each other (Kelty and Landecker, 2009; Mackenzie, 2015). Here, we focus on two case studies to examine how such experiments are designed and executed. From the first incarnations of such experiments to present work, much of the research incorporating large locative datasets has focused on extracting the social, temporal and spatial characteristics within the data; modelling and simulating emerging urban trends and movements; and predicting the flows, patterns and preferences of journeys within and across cities, and even countries (e.g., Gao et al., 2015; Padmanabhan et al., 2014; Xiao et al., 2014). The two studies selected for discussion reflect these tendencies and were selected because they set out their objectives and methods in detail and have had an influence on the wider literature, being well cited or reported. The first section of the paper details how locative data are staged for computing experiments. The second explores how the computing experiments themselves are staged. To make sense of the cases, it is not enough to examine them in isolation. Accordingly, the discussions in the subsequent two sections contextualise the selected computing experiments within broader disciplinary backgrounds before proceeding into analysis. The notion of unintended expectations is elaborated in the next section, which draws broadly upon recent social studies on computationally enhanced, data-driven computing experiments, before concluding the paper.

Staging locative data

The locative data used in computing experiments have their own socioscientific histories of production, appropriation and congregation. Although locative media appeared and started to mature quickly around 2007, the

majority of scientific research that has experimented with locative data has taken place post-2011. The data used in these studies were usually collected over a few hours or weeks, but can incorporate data collected over months or years. The variability is a result of the accessibility to the database, setup of relevant Application Programming Interfaces (APIs), and the methods deployed for data collection, although increasingly the access to the data is merchandised and controlled by large data vendors including GNIP and DataSift. Experiments generally consist of data related to thousands of users and can involve huge numbers of specific actions and locations. For example, experiments using Foursquare data might involve millions of check-ins for hundreds of thousands of locations.

Experiments can be very diverse in nature, differing with respect to the research questions asked, the methods (mathematically, statistically, methodologically and instrumentally) used, the outputs expected and the wider implications the research is expected to have. How they dovetail with existing research interests also varies, as does how these interests motivated the subsequent rendering of data. Across all these studies, however, is a sense of excitement at the availability of such a rich source of new data that seemingly can be repurposed in many ways: 'Data about the interplay between users and locations are for the first time available to researchers, providing unprecedented chances to understand how users actively engage with places and online friends' (Scellato and Mascolo, 2011: 1).

To make locative media data usable, however, they need to be staged, that is cleaned, processed, explored and manipulated to render them fit for repurposing. Just as locative media users 'domesticate' new technology (Silverstone, 1994), researchers have to domesticate locative data by relating their own research interests with the data and translating excitement and uncertainty around the data into actionable expectations. This mostly takes the form of exploratory data analysis, examining what the data reveal about human movement and social ties in particular places.

Observations on check-in data have largely focused on the interplay between temporal, spatial and social factors. For example, Noulas et al. (2011) observed the coupling of locational and temporal dimensions of check-ins and uncovered different patterns of social activities during the week and weekends. They also identified the procession of activities during different periods of time and examined how likely it is for smartphone users to move from one location and activity to the next between two check-ins. The influence of social ties on patterns of check-ins has been another important dimension for the interested researchers to imagine how such data can be reappropriated for predicting future urban movements. In this regard, Cho et al. (2011) observed that people are more likely to travel

a long distance if there is a friend nearby, while short distance travel is less influenced by existing social networks. However, a consistent pattern of the relationships between social ties and check-ins is more difficult to identify than it is to suggest the regularity in temporal patterns of check-ins. As Cho et al. (2011) discussed, even if it is more likely for users having similar trajectories to be friends (at least online), only a small number of users share overlapping check-ins with their friends.

Recognising these inconsistencies, there could be various ways to further unpack the relationships and make them useful for particular purposes. An exploration into the roles, reasons and practices of performing check-ins can be beneficial for designing a location-aware service by leveraging the norms and playfulness facilitated by check-in as meaningful social practices (Cramer et al., 2011). But deciding which factors are important and providing them with appropriate weightings motivate another approach for designing future urban movements. For research concerned with the prediction of preferences in shops or events (e.g., Daly and Geyer, 2011; Gao et al., 2013), it becomes important to find ways to utilise what was known in the existing locative data to infer what was unknown or yet to be known in future events by considering the relationships between spatiality and sociality. This kind of knowledge could be particularly useful for practitioners in the fields of designing mobile advertisements, online service provision, business models, transportation systems, logistics, routing or even epidemics prevention.

To this end, attempts have been made to measure the importance of social and spatial 'structures' in locative media and determine appropriate models for predicting urban social interactions, which can be exemplified in the work by Scellato et al. (2011). They identified correlations between social and spatial structures, followed by a further investigation as to whether it is distance or social relations that could better predict the existence of friendships among existing users. Their initial finding reported that social connections between friends are less likely when distance among them increases. However, knowing that there is positive or negative correlation between social relations and distance was not enough to provide useful predictions regarding how one might conduct inter- and intra-city journeys in relation to where friends are located.

Therefore, a further step was to manipulate and create new datasets to experiment with null models to measure the weight of the impact from social and spatial relations among locative media users. They ran tests on two null models – geo and social – which were created by retaining either social or spatial properties in their original datasets and randomising the other.

For the geo (null) model they kept the information on user location intact, and for social (null) model location information was randomised while data about social connections remained untouched. In doing so, they hoped to observe whether 'socio-spatial characteristics might be explained in terms of simple geographic or social factors' (Scellato et al., 2011: 333). Furthermore, the reprocessing of data can bear the hope that, by comparing properties of new and old datasets after several tests, the influence of either social relation or distance can be measured, compared, weighted and appropriated for when certain data are unavailable or unknown, or in the future when events are unfolding.

Several further tests were made possible by reprocessing the data and creating null models, even though inconclusive results were returned. They tested if the average distance between users and their friends within the null models correspond to the original datasets; if the average distance between individual users and their friends increases with the number of friends and the likelihood that a particular social tie within a larger social network is influenced by distance. The reprocessed datasets and model could not provide predictions that confirm each other. Some tests suggested that the social model might explain the relationships between social ties and space better when, for example, results showed that social relations can appear regardless of the distance between the individuals involved (Scellato et al., 2011: 334). There were tests that seemed to suggest distance played a role in the relationship between sociality and spatiality. The researchers found that at least 20% of users possessed an averaging triangle length of less than 100 km, while the top 20% of users had an average of over 2000 km. Furthermore, when trying to make a prediction concerning the average distance among friends by the number of friends a user has, an increase of distance can be observed when the number of contacts increases. However, the geo model predicts low numbers of contacts among the users whose friends live further away from each other, while the number of contacts a person possesses has no influence on predicting the distance in the social model.

Given these inconsistencies in experiment results, it is not surprising that they suggest with regard to locative datasets: 'their socio-spatial structure cannot be explained by taking into account only geographic factors or social mechanisms' (Scellato et al., 2011: 335). Various trends or correlations observable in the original (real) datasets became insignificant or went missing in the results from testing the null models. A recommendation for refining prediction results was to improve the gravity model by taking into account 'triadic closure' as developed by sociologist Mark Granovetter (1973). The gravity model (Zipf,

1946) and its variations were suggested because it is a long established method for modelling how a user moves from A to B, but adapting the model is considered ‘only a first tentative step’ (Scellato et al., 2011: 336). As the authors recognise, there are several reasons why a gravity model could fail since there are social aspects that are currently not captured in the model and are also difficult to incorporate.

As observed in this section, locative media data are not readymade or purposefully crafted materials. They possess histories, motivations and sociomaterial contexts of production, which become constraints and challenges for computing experiments, resulting in data staging before they can be incorporated and energise existing interests. Many of the tests were run in part to draw insights from the collected datasets, but also to plan next steps of processing and preparing the data for subsequent research developments. Therefore, the rendering of locative data and transforming them into objects that can be experimented with are important steps to sustain the initial momentum created by the emergence of the data and to materialise the excitement around them. What is also observed in staging locative data is that ‘staging’ itself is a technique and computing experiments require staging as much as data processing does, which the following section discusses in more detail.

Staging computing experiments

Just as data are staged to be used in computing experiments, these experiments are themselves staged. By tracing the stepwise procedures implemented in locative media computing experiments it is possible to see how they are setup in relation to the expectations or imagined futures associated with the data. These expectations are partially related to the social and spatial conditions that produce the data, as well as selectively aligned to the interests of the researchers.

To understand the spatial and temporal relationships between check-ins and friendships, for example, Cho et al. (2011: 1086) assumed that ‘certain types of locations, such as home and work, are visited regularly, and often during the same times of the day’. Many issues surface with such a statement. For one, there are complicated social and material orchestrations of networks, geographies and technologies to enable different spatial and temporal configuration of social connections (Larsen et al., 2006), thus highlighting how regularity glosses over nuanced and situated practices of performing urban lives. For another, check-ins can be considered as users’ ongoing experiment with locative media and places (Gordon and de Souza e Silva, 2011). However, this intuitive, highly abstracted observation facilitated the further setting up of their

experiment. As such, Cho et al. (2011) rendered cities into ‘a small set of latent states (locations)’ between which people move. They argue that by focusing on the check-ins around two latent states, representing work and home, urban mobility can be modelled and understood more fully by their modelling, because ‘our model can handle an arbitrary number of them [latent states]’ (Cho et al., 2011: 1086). Here, intuition and commonsensical understandings of the city and instinctive reasoning about urban complexities are drawn upon to inform the design of computing experiments, and the testing of algorithms, in order to produce a ‘successful’ simulation.

Sadilek et al.’s (2013) experiments to simulate urban movement are exemplary of the process whereby intuition was actively used during the design of experimentation and in relating the use of locative media to potential future lives. The motivation of the experiment was to test the feasibility of delivering physical objects by mobilising a crowd (Twitter users) and to demonstrate not only a new use case but also (and more importantly) a multitude of potential and exciting expectations based on their initial work. In other words, the currently unknown (or very little known) future scenarios involving moving physical objects can be rendered through their proposed approach in a way that acts ‘as a motivating example for research on physical crowdsourcing and intelligent coordination of the crowd’ (Sadilek et al., 2013: 2). The benefit of such effort, they claim, is manifold. For one, there is the prospect that a person ‘never has to deviate from her normal route to pick up her package. Instead, it is sent via a chain of people – an algorithm calculates the fastest route using aggregated location data from New York tweeters.’⁶ They also speculate on the ‘initial scenario’ being implemented ‘in poor countries’ with the purpose of rethinking ‘the distribution of vaccines’.

In the experiment, they tested whether the crowd would be a feasible alternative for physical object delivery in two metropolitan cities, New York and Seattle. To do so, they mined geo-tagged tweets and ran simulations to observe whether the tweets, representing the urban movement of respective Twitter users, would likely meet and be able to form a chain of delivery. For the simulation, Seattle was divided into 450 m by 450 m cells for examining if a package can be delivered from a departing cell to a destination cell. A successful delivery is defined in terms of two participants sharing the same locale measured thus: ‘two users meet if they tweet within specified distance and time thresholds’ (Sadilek et al., 2013: 4). Their ambition was to evaluate and improve the effectiveness of routing algorithms in delivering an object when all potential senders and receivers are on the move. This is a significant challenge because previous graph theories presuppose that all

nodes of a graph are known and fixed in the calculation of an optimised route. This assumption cannot be applied when the next leg in a delivery chain needs to be decided ‘on the fly’ so as to allow participants to send and receive while they travel in the city.

To proceed with the research, existing algorithms have to be improved and made fit for the purposes of facilitating the crowd delivery service by taking into account the existing social interaction patterns within the dataset. Accordingly, the locative data provide a potential means to tackle a wicked and unresolved fundamental mathematical problem in terms of responding to the uncertainty in routing when the graph is incomplete or infinite. The ‘local opportunistic algorithm’ Sadilek et al. (2013: 5) proposed takes into account the frequencies of participations appearing near to the parcel’s destination and meeting with each other. It was then compared with two existing algorithms: random algorithm (randomly choosing the next participant to complete the delivery) and global optimum algorithm (shortest paths between all pairs of cells) to explore if they gained efficiency over random selection of participants and gained flexibility by mobilising the crowd for parcel delivery.

To further stage a complicated computing experiment, a ‘strict’ definition on a ‘successful’ delivery is imposed to provide a clear indication that a chain of delivery is formed and to test the potential of the proposed system. Moreover, the motivation of the participants was suppressed. In the tests, they did not include the possibility that a messenger would willingly travel further from their planned destination for an incentive (e.g., cash payment for passing on the parcel) or would do so as an altruistic act. They also did not include a scenario where a package could be temporarily left at a storage facility in the city to be picked up by another volunteer when the person becomes available for the task, even though it would likely increase the rates of success (Sadilek et al., 2013: 6).

Throughout the experiment, stepwise tests were run to examine various aspects related to their proposed delivery method. The experiment measured the geographic extent to which the cities selected could potentially be covered by the service, as well as the time required for accomplishing such delivery. Further, the issue concerning who were the more crucial participants from the ‘crowd’ for the delivery service was examined by simulation. Their initial test results showed that packages were deliverable if two participants could ‘meet’ within a 200m distance to each other and could wait for 90 min to meet and the proposed service could cover more densely populated areas, that is cells having more than 10 tweets in them in the dataset (and

a 400 m distance and 30 min wait time produce very close effects). Their tests also supported their approach in terms of geographic coverage in that, by allowing an 800m digression and 90 min wait time, over 80% of Seattle could be covered by the proposed service – assuming, of course, that people would trust a chain of strangers to pass their parcel across a city, and these strangers would deviate from their route and wait to pass on the parcel. Clearly such an assumption is invalid, but nonetheless locative data provide a means to construct simulation experiments using real-world movement traces. And with more data becoming available from users volunteering their detailed socio-geographic information on locative media, there is a growing sense that urban futures can be more quickly and precisely captured, modelled, predicted, scaled-up and applied across domains by staging both locative data and computing experiments.

Expectations, uncertainties and futures

These data-driven experiments are explorations of possible future computational urban practices and the mechanisms to act upon the uncertainties that could arise as a result. ‘Unintended expectations’ are unavoidable in such data-driven experimentation, with both the data and computing techniques having to be mended, modified and staged to enable experimentation to occur. As a consequence, the effective assembly of data and computing techniques do not unequivocally translate into innovations in urban services. The expectations for new urban practices and futures through these computing experiments, instead, result in an ‘affectionate assembly’ that deepens, lengthens and amplifies the passions for what these experiments hope rather than expect to achieve. What becomes emphasised through these experiments is the immediacy in responding to the possibilities opened up by newly acquired and assembled datasets to create new computing techniques and future scenarios. However, these experiments also create gaps and slippages through their diverse temporalities and sociomaterial practices in producing data, techniques, knowledges and hopes for alternative futures. Accordingly, while these computing experiments seem to offer new avenues and opportunities for innovating data-driven cities, they also create fissures with respect to urban rhythms, practices and patterns that might not be fully discernable. This section draws more broadly upon recent social scientific studies on data-intensive experiments and practices to elaborate the consequences of unintended expectations and reflect upon the ways in which the interests in the city are transformed into passions for computing techniques in the two cases discussed above.

The excitement of witnessing and manipulating locative data emerges alongside the prospect of modelling, visualising, testing and manufacturing unknown futures. Techniques and experiments that are developed to incorporate locative data grow their importance because they provide a material means of 'preemptive' imagining of unknown future cities (Massumi, 2007). These computing experiments operate in the condition that both the future of cities and locative data are indeterminate, and they provide a means to act on that uncertainty by generating diverse methods, simulations and predictions that improve previous experiments and seek to further actualise an unknown future.

To be sure, data are not the only way to respond to the indeterminacy when societies are faced with new possibilities, uncertainties or threats (Adey, 2009). However, echoing Amoore's observation (2011), the techniques of identifying associations among different elements in locative data and the operative logic of preemption materialised in the staging of computing experiments provide an alternative way of envisioning new services for the city. These techniques are novel forms of urban mechanisms that shift the focus away from delineating complicated relationships behind urban problems. Instead, it highlights computational capability of solving the problems by establishing associations from large, disparate, disaggregated and integrated data and transforming uncertain futures into something immediately actionable. These techniques thus produce urban futures by dissecting urban everyday practices into various parcels of data, as well as arraying and flagging differentiated feasibilities (see also Amoore and Hall, 2009). Accordingly, the associations identified in these datasets are the 'excess', what is becoming, escapes and goes on (Anderson, 2010), emerging from data-driven preemptive measures for mitigating, acting upon and governing unknown futures of any city.

More importantly, in order for the proposed futures to be valid, feasible and persuasive, these computing experiments functioning as alternative mechanisms of envisioning urban futures are also tested, seeking to establish its validity and reliability. While it can be argued that science has long been assembling disparate data, methods, technologies and approaches together to seek new synergies and create new sociotechnical futures, the processes of achieving them are far from linear, stable, progressive and ideologically neutral. Expectations, alongside hopes, promises or visions, which come into play in shaping scientific knowledges and technological futures, are entangled in performative processes that enact and normalise certain wishes and desirable futures, as well as managing risks and fears associated with the change. As Borup et al. (2006) argue, the 'value' of technology is difficult to

separate from the intensity of passions and interests mobilised by relevant industries. The expectations around technological innovations are further contextualised by various discourses, knowledges, practitioners, users and practices and materialised in their own histories of contestation, contingency and unfulfilled promises. Furthermore, discourses about scientific achievements are as pervasive as the moments of failed futures in the past, and therefore highlighting moments of breakthrough glosses over the extended histories that led to successful experiments, and misrepresenting drawbacks and contingencies present in scientific and technological developments (Brown and Michael, 2003). In other words, futures are not always developed with linearity. Instead, innovative concepts, experiments, practices and hopes are situated in the social and material conditions of the past that prefigure contemporary expectations associated with sciences (Brown et al., 2006; Tutton, 2012).

Following their work, the staging of locative data and computing experiments cannot be separated from the complex temporalities in which expectations, uncertainties, knowledge practices, everyday routines and contingencies, and technological hopes, promises and innovations relate and reshape one another. These data-driven experiments possess not only an ontology of association (Amoore, 2011), but also techniques of temporalities that dissect and reassemble different paces, rhythms, rates and practices of generating new data, knowledges and urban services. The staging of techniques, as exemplified in the second case study, is filled with intuition, common sense observations, assumptions, tweaks and the playing with parameters, with less regard to the motivations that encouraged the check-ins in the first place or the social and material incentives that might change how these check-ins are performed. Further, making an experiment work takes over the importance of drawing hypotheses from theories and testing them from carefully sampled data that characterises the 'obsolete' sense of science (Anderson, 2008). However, contrary to being free from any theory and knowledge when discovering correlation, as some would claim (e.g., Steadman, 2013), these experiments display 'technics of relation' (Fuller and Goffey, 2012) that has tight control over the assembly of tests and simulations to uncover or formulate particular relations within the large and sparse datasets mined from locative media. These computing experiments demonstrate the mechanics of the 'empiricist epistemology' of data analytics (Kitchin, 2014), which heavily rely on the highly unstable and volatile but attentively crafted steps of experimentation and justification of modelling results. In the experiments, how success is determined and how it was observed in the simulation are heavily dependent upon the deployment of a selective

imagination and tight controls over how people move, why they move, how they report their movements, for what purposes they do so and, most importantly, how they would react to different motivations and incentives. The exclusion of motivation in participating in the chain of delivery, and the exploration of key participants in the chain, are then key examples of how experiments can be run with varying degrees of success, as well as how urban lives are repurposed according to such techniques. Accordingly, the 'potential' of the proposed delivery mechanism does not rest on the feasibility of the service, but on expecting that the success rates of the techniques developed in the experiment would warrant another incarnation of repurposing urban rhythms.

Similarly, the locative data in the first case study are staged to dissect and repackage diverse temporalities associated with producing check-ins for the purpose of experimentation. The use of null models to test and verify the relationship between sociality and spatiality might be a peculiar method, but it shows in great detail that locative data were purposefully staged in order to fit the data into the expected research purposes and procedures. Locative media users do not check-in to stores for producing systematic entries to record their life history. From the perspective of locative data, they possess spontaneity, sparsity and improvisation, and are products of the diverse ways in which users perform identity (Cramer et al., 2011), conspicuous consumption (Wilson, 2012), 'micro-coordination' by check-ins and personal relationships that are attached to physical and virtual places (Ciolfi and Avram, 2016; Ling and Yttri, 2002). Accordingly, these different practices of the self, time, places and personal relationships shape locative data in particular ways and are entangled with and difficult to separate from the spatiotemporal rhythms around geo-tagged places. These somewhat peculiar practices are, however, preserved in the data and demand experiment designs and procedures to be crafted to fit the data into experiments. Particular aspects of the data are amplified or removed, and different datasets are created for the evaluation of the techniques to manipulate and govern the data in relation to the effects specific computing experiments expect to achieve.

Furthermore, techniques to dissect and assemble the paces and practices of performing check-ins are situated in comparatively accelerated procedures and intensified rates of knowledge accumulation and commercial product provision (see Rabinow, 2005). Observing this process in the context of synthetic biology, Mackenzie (2013) argues that different rates of realisation embody gaps, frictions and slippages among techniques developed to assemble biological parts, as well as other follow-up 'techniques' to assemble the composite with

often limiting social, economic, political and industrial environments. Laboratory and engineering techniques are developed and designed in response to certain purposes, standards, building processes, procedures and networks of infrastructure, which in turn modify the techniques or request new ones. It is within these processes, he argues, where unpredictability and the politics of promises and futures are opened up and can be observed and mapped by detailed descriptions.

Accordingly, rates of realisation provide an important lens to reveal the gaps, slippages and consequences as a result of leveraging data-driven experiments for envisioning urban futures. Apart from the urban rhythms of check-ins, there are additional, diverse temporalities packed into the experiments, despite an over-emphasis on the immediacy of predicting urban futures. Data collection is the most rapid one, completed within only a few hours or days. This contrasts with slower processes of experiment design, data wrangling and exploratory data analysis. Although location-based services and location-based social networking both have relatively short histories, it still took over 15 months before computing research started to use check-ins as data, and longer still for such research to become more common. Furthermore, when validating check-in as data for understanding urban mobility, computing research often cites or compares check-in patterns with other research analysing mobile phone signals for extracting urban mobility patterns. However, getting access to a snapshot of mobile phone data is difficult, expensive and time consuming (Ahas et al., 2010). Moreover, captured in the locative data and confirmed by the experiments are urban rhythms, such as working hours, commuting patterns, week days and weekends, that are enabled by the infrastructure of experiences (Dourish and Bell, 2007) and mundane but critical work of maintenance and repair to performing everyday tasks (Perng, 2015). These are products of long histories of sociotechnical inventions, such as clocks and timetables, starting even before the industrialisation of urban rhythms (Glennie and Thrift, 2009). As a result, what are often predicted by the experiments are fine-tuned urban rhythms shaped and embedded in prolonged sociotechnical processes, rather than emerging trends or new practices of the immediate future.

Accordingly, what these data-driven experiments achieve is less about the predictions or preemptions of uncertain urban futures than it is about an affective mobilisation of data techniques as an accelerated alternative. The staging of data and computing experiments is an affective 'imitation' that animates and sustains the ongoing and continuous efforts of transforming urban future into simulation and prediction (Tarde, 1903). Data derivatives and computing experiments are both a means of bringing the future to the present and

rendering uncertainty actionable. However, these experiments do not produce immediately actionable future. Rather, what is enacted is that computing experiments become ‘a machine for promoting *passionate imitation*’ (Barry and Thrift, 2007; original emphasis). For Tarde, imitation is central to the process of innovation in that inventive activities both repeat and improve the techniques and knowledge that are already established, and make them better fitted for other situations, goals or more fundamentally desires (Tarde, 1903: 94). Invention thus produces differences and repetitions, as well as momentum that extends and deepens the passionate interests that continue to refine previous efforts. Understood this way, crafting computing techniques to render locative data actionable is yet another inventive activity that produces and registers differences, repetitions and contagion, alongside archaeology, economics and statistics that Tarde analyses.

Therefore, through staging locative data and computing techniques, data-driven experiments simultaneously ‘flatten’ and ‘thicken’ certain worlds, dependent upon the techniques invented ‘to track or extract forms of identity, regularity or pattern that we are not able to see, make or say directly or immediately’ (Mackenzie and McNally, 2013: 73–75). To render unknown urban future actionable, locative data have to be processed, manipulated and fitted into specific sets of assumptions, procedures, methods and tests. In parallel, theories, models, creativity, intuition, knowledge and understanding of the past are mobilised, modified and merged to respond to check-ins and the unknown futures that such data might trigger. Success, however, is not entirely the imaginations, visions and promises made possible by the experiments. Instead, it is a momentum associated with, and promoted by, a multitude of techniques for staging data that are materialised in the process. It is an affective stabilisation of multiple pathways and possibilities for the mutual fitting of data and techniques that enthruses the fast uptake of the locative data and fast-track a sub-discipline (city science or multiple sub-disciplines if considering diverse facets and applications of ‘data science’). And yet, these experiments reduce temporal depth and complexity, and result in proposing versions of the future that are only remotely realisable. Computing experiments therefore make important sites of examination to explore how an unknown future has been rendered and processed, and how expectations have been sensed, expressed and actualised (c.f. Massumi, 2002: 34–37).

Conclusion

In this paper we have sought to widen the remit of critical data studies to consider the unanticipated uses

of data and the unintended consequences that arise from such usage. Accordingly, we have examined how locative media data have become the stimulus for new computing experimentation with the data repurposed to try and solve diverse problems. In such experiments, locative media data are being reimagined and reworked through a process of staging that unfolds in contingent, relational and unexpected ways through competing discourses, knowledges and practices which produce all kinds of unintended consequences. In short, our aim has been to move beyond questioning the production, epistemologies, ontologies and uses of data (e.g., boyd and Crawford, 2012; Kitchin, 2014; Wilson, 2015), by demonstrating new ways of interrogating and conceptualising how data are staged for various forms of processing and analysis and how the mechanisms of processing and analysis are themselves staged to interface with the data. This approach, we believe, holds value because it challenges critical data scholars to think carefully about how the same dataset is rethought and reworked in diverse and unexpected ways, and has multiple parallel and intersecting lives and can mutate into multiple incarnations.

As we have illustrated, from Foursquare to smart city initiatives, there is now a growing interest in making better predictions about urban futures by repurposing large datasets, generated by ordinary users. One of the key pathways to such urban futures is computing experiments that seek to push the boundaries of conventional thinking and find novel solutions through structured playing with new datasets. The main perceived advantage of such data is that they are generated ‘in the wild’, outside of laboratories and research settings, and are recordings of real trajectories and interactions. However, as our analysis demonstrates, significant challenges arise when repurposing locative data, and therefore the practices of staging and the conflated expectations and interests in computing experiments require further attention. Locative data were not produced for testing existing epistemological and knowledge assumptions and gaps in the modelling and prediction of urban practices. As such, new statistical models and computing techniques have to be innovated to make sense of, manipulate, utilise, derive, imagine and use these data to create new knowledges about present and possible future cities. In some cases, it becomes clear that access to such a rich new dataset and the perceived possibilities concerning its analysis are driving the science, rather than the science being driven by solving a specific question.

Set against this backdrop, this paper has proposed the notion of unintended expectations to capture and analyse the emergence and effect of staging data and computing experiments. Recognising that unintended consequences have become a significant aspect of

considering the relation between technology and society, the paper has taken the research upstream to argue that, particularly under the context of data-driven computing work, expectations can be inadequately set in the first place and therefore various steps and interventions have to be put in place to make data and computing techniques fit for each other, as well as for purposes for which they are not intended. And yet, because of the dramatically accelerated rate of realising computing experiments and verifying scenarios about urban futures, momentum is created and passion for the experimental city deepened.

It is open to debate, or to experiment, whether the prediction or simulation of sociospatial interactions or parcel delivery in the city is successful. But what this paper sensitises and proposes for further critical examination are the tweaks, modifications and improvements, such as those observed above, that enable and intensify the interests of pursuing yet another experiment to test updated or renewed scenarios about future cities. Particularly because locative data are produced for purposes other than experiments, they are not intangible, easily malleable sets of 0s and 1s. Instead, they are objects that require techniques, creativity, innovative methods and imaginations to start interacting with them and shaping expectations from them. What is promoted and energised throughout the process, however, is not simply the power of prediction offered by techniques of modelling and algorithms. They play a part, but more crucially, what is promised is an affectionate assembly of data and techniques in the name of generating new imaginations for future cities. Such mobilisation is achieved by rendering uncertainty into something knowable, and further transforming it into computing experiments that can be run, tested, validated, refuted but most importantly further imitated.

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research for this paper was funded by a European Research Council Advanced Investigator Award, 'The Programmable City' (ERC-2012-AdG-323636).

Notes

1. Both launched in 2007 but were acquired by Limbo in 2009 and Facebook in 2012, respectively.

2. Details in <http://expandedramblings.com/index.php/resource-how-many-people-use-the-top-social-media/4/#.Uz0qGo8julM> (accessed 3 April 2014).
3. Details in <http://www.statisticbrain.com/facebook-statistics/> (accessed 3 April 2014).
4. Details in <http://expandedramblings.com/index.php/by-the-numbers-interesting-foursquare-user-stats/#.Uz0iB48julM> (accessed 3 April 2014).
5. Snapchat Privacy Policy: <https://www.snapchat.com/privacy> (accessed 6 July 2015).
6. Source of this and the next quotes: <http://www.newscientist.com/article/mg21829175.500-parcels-find-their-way-to-you-via-the-crowd.html#.Uz1CrI8julM> (accessed 19 March 2014).

References

- Adey P (2009) Facing airport security: Affect, biopolitics, and the preemptive securitisation of the mobile body. *Environment and Planning D: Society and Space* 27(2): 274–295.
- Ahas R, Silm S, Järvi O, et al. (2010) Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of Urban Technology* 17(1): 3–27.
- Amoore L (2011) Data derivatives on the emergence of a security risk calculus for our times. *Theory, Culture and Society* 28(6): 24–43.
- Amoore L and Hall A (2009) Taking people apart: Digitised dissection and the body at the border. *Environment and Planning D: Society and Space* 27(3): 444–464.
- Anderson B (2010) Preemption, precaution, preparedness: Anticipatory action and future geographies. *Progress in Human Geography* 34(6): 777–798.
- Anderson C (2008) The end of theory: The data deluge makes the scientific method obsolete. *Wired*, 23 June 2008. Available at: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory (accessed 15 September 2015).
- Barnes TJ and Wilson MW (2014) Big Data, social physics, and spatial analysis: The early years. *Big Data and Society* 1(1): 1–14.
- Barreneche C and Wilken R (2015) Platform specificity and the politics of location data extraction. *European Journal of Cultural Studies* 18(4–5): 497–513.
- Barry A and Thrift N (2007) Gabriel Tarde: Imitation, invention and economy. *Economy and Society* 36(4): 509–525.
- Borup M, Brown N, Konrad K, et al. (2006) The sociology of expectations in science and technology. *Technology Analysis and Strategic Management* 18(3–4): 285–298.
- boyd d and Crawford K (2012) Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society* 15(5): 662–679.
- Brown N, Kraft A and Martin P (2006) The promissory pasts of blood stem cells. *BioSocieties* 1(3): 329–348.
- Brown N and Michael M (2003) A Sociology of expectations: Retrospecting prospects and prospecting retrospects. *Technology Analysis and Strategic Management* 15(1): 3–18.
- Cho E, Myers SA and Leskovec J (2011) Friendship and mobility: User movement in location-based social

- networks. In: *Proceedings of the 17th international conference on knowledge discovery and data mining*, San Diego, CA, USA, 21–24 August 2011, pp. 1082–1090. NY: ACM.
- Ciolfi L and Avram G (2016) Digital social interactions in the city: Reflecting on location-based social networks. In: Kitchin R and Perng S-Y (eds) *Code and the City*. London: Routledge, pp. 91–104.
- Cramer H, Rost M and Holmquist LE (2011) Performing a check-in: Emerging practices, norms and ‘conflicts’ in location-sharing using Foursquare. In: *Proceedings of the international conference on human computer interaction with mobile devices and services*, Stockholm, Sweden, 30 August–2 September 2011, pp. 57–66. NY: ACM.
- Dalton C and Thatcher J (2014) What does a critical data studies look like, and why do we care? Seven points for a critical approach to ‘big data’. *Society and Space Open Site: Run by the Editors of the Journal Environment and Planning D: Society and Space*. Available at: <http://societyandspace.com/material/commentaries/craig-dalton-and-jim-thatcher-what-does-a-critical-data-studies-look-like-and-why-do-we-care-seven-points-for-a-critical-approach-to-big-data/> (accessed 1 February 2016).
- Daly EM and Geyer W (2011) Effective event discovery: Using location and social information for scoping event recommendations. In: *Proceedings of the fifth conference on recommender systems*, Chicago, IL, USA, 23–27 October 2011, pp. 277–280. NY: ACM.
- Dourish P and Bell G (2007) The infrastructure of experience and the experience of infrastructure: Meaning and structure in everyday encounters with space. *Environment and Planning B: Planning and Design* 34(3): 414–430.
- Edwards L (2015) Privacy, security and data protection in smart cities: A critical EU law perspective. CREATE Working Paper 2015/11. Available at: <https://zenodo.org/record/34501/files/CREATE-Working-Paper-2015-11.pdf> (accessed 1 February 2016).
- Evans L (2013) How to build a map for nothing: Immaterial labour and location-based social networking. In: Lovink G and Rasch M (eds) *Unlike Us Reader: Social Media Monopolies and Their Alternatives*. Amsterdam: Institute of Network Cultures, pp. 189–199.
- Evans L (2015) *Locative Social Media: Place in the Digital Age*. London: Palgrave Macmillan.
- Fuller M and Goffey A (2012) Digital infrastructures and the machinery of topological abstraction. *Theory, Culture and Society* 29(4–5): 311–333.
- Gao H, Tang J, Hu X, et al. (2013) Exploring temporal effects for location recommendation on location-based social networks. In: *Proceedings of the 7th conference on recommender systems*, Hong Kong, 12–16 October 2013, pp. 93–100. NY: ACM.
- Gao H, Tang J, Hu X, et al. (2015) Content-aware point of interest recommendation on location-based social networks. In: *Proceedings of the 29th AAAI conference on artificial intelligence*, Austin, TX, USA, 25–29 January 2015, pp. 1721–1727. Palo Alto, California: The AAAI Press.
- Glennie P and Thrift N (2009) *Shaping the Day: A History of Timekeeping in England and Wales 1300–1800*. Oxford: Oxford University Press.
- Gordon E and de Souza e Silva A (2011) *Net Locality: Why Location Matters in a Networked World*. Chichester, Sussex: Wiley-Blackwell.
- Graham M and Zook M (2013) Augmented realities and uneven geographies: Exploring the geolinguistic contours of the web. *Environment and Planning A* 45(1): 77–99.
- Granovetter M (1973) The strength of weak ties. *American Journal of Sociology* 78(6): 1360–1380.
- Kelty C and Landecker H (2009) Ten thousand journal articles later: Ethnography of ‘The literature’ in science. *Empiria. Revista de metodología de ciencias sociales* 18, Julio–Diciembre: 173–192.
- Kinsley S (2011) Anticipating ubiquitous computing: Logics to forecast technological futures. *Geoforum* 42(2): 231–240.
- Kitchin R (2014) Big Data, new epistemologies and paradigm shifts. *Big Data and Society* 1(1): 1–12.
- Kitchin R (2015) Data-driven, networked urbanism. Social Science Research Network. Available at: <http://papers.ssrn.com/abstract=2641802> (accessed 26 September 2015).
- Kitchin R (2016) *Getting Smarter About Smart Cities: Improving Data Privacy and Data Security*. Dublin, Ireland: Department of the Taoiseach, Data Protection Unit. Available at: http://www.taoiseach.gov.ie/eng/Publications/Publications_2016/Smart_Cities_Report_January_2016.pdf (accessed 1 February 2016).
- Kitchin R and Lauriault T (2014) Towards critical data studies: Charting and unpacking data assemblages and their work. The Programmable City Working Paper 2, SSRN. Available at: <http://ssrn.com/abstract=2474112> (accessed 1 February 2016).
- Larsen J, Urry J and Axhausen KW (2006) *Mobilities, Networks, Geographies*. Aldershot: Ashgate.
- Latour B and Lépinay VA (2009) *The Science of Passionate Interests: An Introduction to Gabriel Tarde’s Economic Anthropology*. Paradigm, Chicago: Prickly Paradigm Press.
- Leonelli S (2014) What difference does quantity make? On the epistemology of Big Data in biology. *Big Data and Society* 1(1): 1–11.
- Ling R and Yttri B (2002) Hyper-coordination via mobile phones in Norway. In: Katz E and Aakhus MA (eds) *Perpetual Contact: Mobile Communication, Private Talk, Public Performance*. Cambridge: Cambridge University Press, pp. 139–169.
- McArdle G, Furey E, Lawlor A, et al. (2014) Using digital footprints for a city-scale traffic simulation. *ACM Transactions on Intelligent Systems and Technology* 5(3): 41:1–41:16.
- Mackenzie A (2013) Realizing the promise of biotechnology: Infrastructural-icons in synthetic biology. *Futures* 48: 5–12.
- Mackenzie A (2015) The production of prediction: What does machine learning want? *European Journal of Cultural Studies* 18(4–5): 429–445.
- Mackenzie A and McNally R (2013) Living multiples: How large-scale scientific data-mining pursues identity and differences. *Theory, Culture and Society* 30(4): 72–91.
- Massumi B (2002) *Parables for the Virtual: Movement, Affect, Sensation*. Durham, NC: Duke University Press.

- Massumi B (2007) Potential politics and the primacy of pre-emption. *Theory and Event* 10(2). Available at: http://muse.jhu.edu/journals/theory_and_event/v010/10.2massumi.html (accessed 21 May 2014).
- Noulas A, Scellato S, Mascolo C, et al. (2011) An empirical study of geographic user activity patterns in Foursquare. In: *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain, 17–21 July 2011, pp. 570–573. Menlo Park, CA: The AAAI Press.
- Padmanabhan A, Wang S, Cao G, et al. (2014) FluMapper: A cyberGIS application for interactive analysis of massive location-based social media. *Currency and Computation: Practice and Experience* 26(13): 2253–2265.
- Perng S-Y (2015) Performing tasks with wi-fi signals in Taipei. *Space and Culture* 18(3): 285–297.
- Rabinow P (2005) *A Machine to Make a Future: Biotech Chronicles*. Princeton, NJ: Princeton University Press.
- Sadilek A, Krumm J and Horvitz E (2013) Crowd physics: Planned and opportunistic crowdsourcing for physical tasks. In: *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, Cambridge, MA, USA, 8–11 July 2013, pp. 536–545. Palo Alto, CA: The AAAI Press.
- Scellato S and Mascolo C (2011) Measuring user activity on an online location-based social network. In: *IEEE conference on computer communications workshops*, Shanghai, China, 10–15 April 2011, pp. 918–923. Piscataway, NJ: IEEE.
- Scellato S, Noulas A, Lambiotte R, et al. (2011) Socio-spatial properties of online location-based social networks. In: *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain, 17–21 July 2011, pp. 918–923. Menlo Park, CA: The AAAI Press.
- Silverstone R (1994) *Television and Everyday Life*. London: Routledge.
- Steadman I (2013) Big data and the death of the theorist. *Wired*, 25 January 2013. Available at: <http://www.wired.co.uk/news/archive/2013-01/25/big-data-end-of-theory> (accessed 25 September 2015).
- Tarde G (1903) *The Laws of Imitation*. New York: Holt and Company.
- Tarde G (2007) Economic psychology. *Economy and Society* 36(4): 614–643.
- Tutton R (2012) Personalizing medicine: Futures present and past. *Social Science and Medicine* (1982) 75(10): 1721–1728.
- Wilson MW (2012) Location-based services, conspicuous mobility, and the location-aware future. *Geoforum* 43(6): 1266–1275.
- Wilson MW (2015) Morgan Freeman is dead and other social media stories. *Cultural Geographies* 22(2): 345–349.
- Xiao X, Zheng Y, Luo Q, et al. (2014) Inferring social ties between users with human location history. *Journal of Ambient Intelligence and Humanized Computing* 5(1): 3–19.
- Zipf GK (1946) The P1 P2/D hypothesis: On the intercity movement of persons. *American Sociological Review* 11(6): 677–686.

This article is a part of Special theme on Critical Data Studies. To see a full list of all articles in this special theme, please click here: <http://bds.sagepub.com/content/critical-data-studies>.