



Dissertation submitted to the Department of Media and Communications, London School of Economics and Political Science, August 2014, in partial fulfilment of the requirements for the MSc in Media, Communication and Development. Supervised by Dr Ellen Helsper.

The Author can be contacted at: cgy.gordon@gmail.com

Published by Media@LSE, London School of Economics and Political Science ('LSE'), Houghton Street, London WC2A 2AE. The LSE is a School of the University of London. It is a Charity and is incorporated in England as a company limited by guarantee under the Companies Act (Reg number 70527).

Copyright in editorial matter, LSE © 2015

Copyright, Charly Gordon © 2015.
The authors have asserted their moral rights.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means without the prior permission in writing of the publisher nor be issued to the public or circulated in any form of binding or cover other than that in which it is published. In the interests of providing a free flow of debate, views expressed in this dissertation are not necessarily those of the compilers or the LSE.

Big Data exclusions and disparate impact: investigating the exclusionary dynamics of the Big Data phenomenon

Charly Gordon

ABSTRACT

In the past few years, there have been seemingly endless hopes and claims made about the potential benefits of Big Data for society. The phenomenon has spurred the publication of countless media reports and fostered significant academic research in a variety of fields. However, despite this abundance of literature, the Big Data phenomenon suffers from considerable research gaps. Specifically, there has recently been a growing interest in the social consequences of Big Data with particular attention given to the phenomenon's potential to aggravate structural inequalities.

This study offers an inquiry into the phenomenon's impact in terms of social exclusion and its potential for harmful discrimination. By conducting semi-structured elite interviews, this paper seeks to assess the apparent contradiction between Big Data's dynamics of exclusion and the dominant conceptual projection of Big Don

(3) Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy

(education, age, gender and culture) should be taken into account in addition to ICT

the 'Petabyte Age'

Yet the lack of transparency caused by data monopolization strategies often make it difficult to implement data oversight and adequately assess both the methodology and ethics of the research that is being conducted.

The adverse effects of data monopolization

Because access to data and the ability to draw meaningful inferences from it hold so much value, private actors have an economic incentive to create scarcity through monopolization strategies. Much of the data flows produced everyday are caught by a multitude of private sector organizations, making access to the data arduous or even impossible for outsiders. Policies regarding access to proprietary data vary from one company to another. Many organizations restrict access to their data completely while others may sell it, trade it or on some occasions offer small data sets for research purposes. Nevertheless this data insider-outsider dichotomy 'produces considerable unevenness in the system' (Boyd and Crawford, 2012). This may be problematic for three reasons. First, limited access to data may lead to a lack of methodological oversight, which is essential given the complexities and intricacies of data creation mechanisms. This lack of methodological oversight can have dire consequences when the sweeping claims made by what Kate Crawford calls the 'Big Data rich' 08001880 45 0441.92cm

Claims of Big Data discrimination

The myth of algorithmic neutrality

Finally, the risks developed previously may be further entrenched by the myth of algorithmic neutrality. With the Big Data phenomenon, many businesses and public actors 'foster an illusion that classification is (or should be) an area of absolute algorithmic rule—that decisions are neutral, organic, and even automatically rendered without human intervention—

Rationale for method used

Why opt for qualitative semi-structured interviews?

The purpose of this study is exploratory in nature. It seeks to shed light on how specific aspects of a phenomenon, that have yet to be fully researched, are assessed by those who hold a privileged position in the field of Big Data. Therefore, my study is inherently qualitative and due to its exploratory nature, a semi-structured interview method was deemed appropriate. One of the strengths of this method is that it gives time and leeway to explore a variety of issues thoroughly. After specific topics in line with my research questions were evoked and discussed, loose, open-ended questions let the interviewees formulate answers with more freedom. This allowed them to drive the discussion towards issues they felt were important while enabling me to discover and tackle new issues (Bauer and Gaskell, 2000). Conducting highly structured interviews would have left less room for serendipity and would therefore have been detrimental to this research project. Furthermore, a face-to-

phenomenon' (Bourdieu, 1991: 170). It is because these 'elite' agents have the ability to construct the reality of the Big Data phenomenon that they are central to this paper.

Particularities of elite interviewing

Before analyzing my sampling and my recruitment method it is important to look at some of the particularities of elite interviewing. Though elite interviews are more prevalent in journalism than in academia, many scholars have written about the intricacies of elite interviewing as a method of research in social sciences (Dexter, 1970; Merton, Fiske and Kendall, 1990; Ostrander, 1995; Odendahl

Methodological caveats

In addition, there was a consensus on the shortcomings of the regulatory framework to tackle both current and future challenges posed by Big Data. The European model posits the protection of '

Furthermore, because of the craze for Big Data in the private sector, 'data fundamentalism' was recognized as having potentially adverse social consequences as organizations

In this context, the question of whether or not certain precautions are taken when data is analyzed by organizations was also addressed. 'One of the problems with Big Data is all the buzz around the phenomenon. Inferring the behaviour of a population by studying tweets is fundamentally biased. Twitter users don't represent Internet users and Internet users don't represent the wider population etc. There is a fundamental skew and certain companies have the feeling that because it can be measured it somehow becomes representative,' warned D.C. In addition to the digital production gap, he believes there is a gap in expertise in organizations to properly draw conclusions from these new sources of data. 'Data analys2 (() 1 1841.9 -2t(a)

system underlined F.M. Therefore, it is difficult to envision a legal solution that reinforces individual control over the use of their data through ownership mechanisms. Furthermore, 'there is a real problem in the valuation of data,' stressed L.T. An individual's data on its own holds very little value. It is only once it can be crossed with multiple sources that it becomes valuable. In addition, individuals are often unaware of which data are being harvested by organizations – 'we don't know what it is exactly we're selling when we're creating data trails,' added F.M. Nevertheless, interviewees acknowledged that regaining consumer sovereignty is paramount to avert the negative consequences linked to digital exclusion and Big Data's disparate impacts.

even the developers', he added. This raises another issue. Many patterns and correlations detected by Big Data analytics may be counterintuitive and making sense of the result may prove difficult. 'Usually the models are in line with our intuition, often with greater precision, but sometimes they go against our intuition. We're sometimes tempted to change our decision because of counterintuitive results... but it's difficult,' recognized D.B. Even though French law requires human intervention when individuals are impacted by automated decision mechanisms, how does one reverse a potentially adverse outcome if it is difficult to

terms of social exclusion and its potential for discriminatory outcomes. Second, it looked to investigate how those who hold a privileged position in the field of Big Data appraise the phenomenon's social impact and furthermore to assess their recommendations on how to tackle Big Data's exclusionary dynamics. To achieve this, semi-structured elite interviews were conducted over a period of a month in Paris. Thus, this study offers a snapshot of how French elite actors in the Big Data field evaluate the exclusionary dynamics of Big Data in a French and European regulatory framework. Nevertheless, given the large quantities of research data acquired, the interviews offered a range of opinions from which it was possible to draw a number of findings.

First, although a number of academics have been warning of the pitfalls of data invisibility and the potential adverse effects of the Big Data gap, many interviewees saw this exclusionary

projects bent on insuring greater awareness and transparency in the data economy. In addition to strengthening Open Data initiatives, the budding field of human-data interaction may offer a promising path to achieve greater individual control over data creation mechanisms and data portability. However, additional research is needed to address HDI's many challenges. These include consumer data visualization and making sense of complex mechanisms as well as the nature of the technical infrastructure and the institutional framework to drive such interactions.

REFERENCES

Eastwood, M, Olofson, C

Jerome, J. (2013) Buying and Selling Privacy: Big Data's Different Burdens and Benefits . Stanford Law Review. 66 (1), p47-

Marwick A. (2010) Status Update: Celebrity, publicity, and self-branding in Web 2.0. Unpublished doctoral dissertation, New York University, New York.

Marwick, A. (2014) How Your Data Are Being Deeply Mined, URL: <http://www.nybooks.com/articles/archives/2014/jan/09/how-your-data-are-being-deeply-mined/> [Last consulted 23rd August 2014].

Mayer-Schönberg, V. and Cukier, K. (2012) Big Data: A Revolution That Will Transform How We Live, Work, and Think. 2nd ed. NY: Eamon Dolan/Houghton Mifflin Harcourt.

Merton, R., Fiske, M. and Kendall, P. (1990) The focused interview: A manual of problems and procedures (2nd ed.). New York: Free Press.

Mierzwinski, E. (2014) Protecting Consumer Privacy and Welfare in the Era of 'E-Scores,' Real-time Big- Data 'Lead-Generation' Practices and other Scoring/Profile Applications, URL: http://www.centerfordigitaldemocracy.org/sites/default/files/AltScoringCommentsFTC18March14PIRG&CDD_0.pdf [Last consulted 10th August 2014].

Mortier R. et al. (2013) Challenges & Opportunities in Human-Data Interaction, URL: <http://www.eecs.qmul.ac.uk/~hamed/papers/hdi2013.pdf> [Last consulted 10th August 2014].

Odendahl, T., and Shaw, A. (2002) Interviewing elites. In J. Gubrium & J. Holstein (Eds.), Handbook of interview research: Context and methodology. Thousand Oaks, CA: Sage. p-299-316.

Ohm, P. (2012) Don't Build a Database of Ruin, URL: <http://blogs.hbr.org/2012/08/dont-build-a-database-of-ruin/> [Last consulted 23rd August 2014].

Ostrander, S. (1995) Surely you're not in this just to be helpful: Access, rapport, and interviews on three studies of elites. In R. Hertz & J. Imber (Eds.), Studying elites using qualitative methods. Thousand Oaks, CA: Sage. p-133-150.

Pariser, E (2012) The Filter Bubble, What the Internet is Hiding From You. London: Penguin.

Podesta, J. (2014) Big Data: Seizing Opportunities, Preserving Values, URL: http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [Last consulted 6th July 2014].

Richards, N and King, J. (2013) Three Paradoxes of Big Data. *Stanford Law Review*. 66 (1), p41-46.

Rodgers, G *et al.* (1995) *Social Exclusion: Rhetoric, Reality, Responses*. Geneva: International Institute for Labour Studies.

APPENDIX A: Topic Guide

In your opinion, what are the major challenges linked to Big Data today in developed countries?

3- Big Data's benefits to consumers and citizens

Can you spell out how Big Data and Big Data analytics can benefit consumers and citizens?

Which conditions must be met for Big Data to benefit society at large?

4- Big Data and social exclusions

In your opinion, is Big Data an inclusive or exclusive phenomenon? Is the Big Data phenomenon today socially empowering?

Do the benefits mentioned previously, that stem from Big Data and Big Data analytics, impact all consumers and citizens in the same way? Is it fair to say that there are winners and losers in Big Data? Why?

Can digital exclusions have an impact on Big Data's social outcomes? Can this entrench existing inequalities? Can it create new ones? If so, how can this be adequately addressed?

Can the categorization of individuals lead to new forms of discrimination? Can behavioural targeting have discriminatory outcomes? If so, how can this be adequately addressed?

Do power asymmetries in the data economy lead to adverse social outcomes?

Electronic MSc Dissertation Series

The Media@LSE Electronic MSc Dissertations Series presents high quality MSc Dissertations which received a mark of 73% and above (Distinction).

Selected dissertations are published electronically as PDF files, subject to review and approval by the Editors.

Authors retain copyright, and publication here does not preclude the subsequent development of the paper for publication elsewhere.
