THE UNIVERSITY OF WARWICK

Library Declaration and Deposit Agreement

1. STUDENT DETAILS

Please complete the following:

Full name: ..ANA GROSS.....

University ID number: ....1290198.....

2. THESIS DEPOSIT

2.1 Under your registration at the University, you are required to deposit your thesis with the University in BOTH hard copy and in digital format. The digital copy should normally be saved as a single pdf file.

2.2 The hard copy will be housed in the University Library. The digital copy will be deposited in the University's Institutional Repository (WRAP). Unless otherwise indicated (see 2.6 below), this will be made immediately openly accessible on the Internet and will be supplied to the British Library to be made available online via its Electronic Theses Online Service (EThOS) service.
[At present, theses submitted for a Master's degree by Research (MA, MSc, LLM, MS or MMedSci) are not being deposited in WRAP and not being made available via EthOS. This may change in future.]

2.3 In exceptional circumstances, the Chair of the Board of Graduate Studies may grant permission for an embargo to be placed on public access to the thesis **in excess of two years**. This must be applied for when submitting the thesis for examination (further information is available in the *Guide to Examinations for Higher Degrees by Research*.)

2.4 If you are depositing a thesis for a Master's degree by Research, the options below only relate to the hard copy thesis.

2.5 If your thesis contains material protected by third party copyright, you should consult with your department, and if appropriate, deposit an abridged hard and/or digital copy thesis.

2.6 Please tick one of the following options for the availability of your thesis (guidance is available in the *Guide to Examinations for Higher Degrees by Research)***:**

- [ ] Both the hard and digital copy thesis can be made publicly available immediately
- [ ] The hard copy thesis can be made publicly available immediately and the digital copy thesis can be made publicly available after a period of two years (*should you subsequently wish to reduce the embargo period please inform the Library*)
- [X] Both the hard and digital copy thesis can be made publicly available after a period of two years (*should you subsequently wish to reduce the embargo period please inform the Library*)
- [ ] Both the hard copy and digital copy thesis can be made publicly available after _____ (insert time period in excess of two years). **This option requires the prior approval of the Chair of the Board of Graduate Studies (see 2.3 above)**

The University encourages users of the Library to utilise theses as much as possible, and unless indicated below users will be able to photocopy your thesis.

- [ ] I **do not** wish for my thesis to be photocopied

3. GRANTING OF NON-EXCLUSIVE RIGHTS

Whether I deposit my Work personally or through an assistant or other agent, I agree to the following:

- Rights granted to the University of Warwick and the British Library and the user of the thesis through this agreement are non-exclusive. I retain all rights in the thesis in its present version or future versions. I agree that the institutional repository administrators and the British Library or their agents may, without changing content, digitise and migrate the thesis to any medium or format for the purpose of future preservation and accessibility.

4. DECLARATIONS

I DECLARE THAT:

- I am the author and owner of the copyright in the thesis and/or I have the authority of the authors and owners of the copyright in the thesis to make this agreement. Reproduction of any part of this thesis for teaching or in academic or other forms of publication is subject to the normal limitations on the use of copyrighted materials and to the proper and full acknowledgement of its source.

- The digital version of the thesis I am supplying is either the same version as the final, hard-bound copy submitted in completion of my degree once any minor corrections have been completed, or is an abridged version (see 2.5 above).

- I have exercised reasonable care to ensure that the thesis is original, and does not to the best of my knowledge break any UK law or other Intellectual Property Right, or contain any confidential material.

- I understand that, through the medium of the Internet, files will be available to automated agents, and may be searched and copied by, for example, text mining and plagiarism detection software.

- At such time that my thesis will be made publically available digitally (see 2.6 above), I grant the University of Warwick and the British Library a licence to make available on the Internet the thesis in digitised format through the Institutional Repository and through the British Library via the EThOS service.

- If my thesis does include any substantial subsidiary material owned by third-party copyright holders, I have sought and obtained permission to include it in any version of my thesis available in digital format and that this permission encompasses the rights that I have granted to the University of Warwick and to the British Library.

5. LEGAL INFRINGEMENTS

I understand that neither the University of Warwick nor the British Library have any obligation to take legal action on behalf of myself, or other rights holders, in the event of infringement of intellectual property rights, breach of contract or of any other right, in the thesis.

Please sign this agreement and ensure it is bound into the final hard bound copy of your thesis, which should be submitted to Student Reception, Senate House.

Student's signature: Ana Gross..... Date: .....24/11/2015.....

Data Types and Functions:
A Study of Framing Devices and Techniques

Ana Gross

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy in Interdisciplinary Studies

Centre For Interdisciplinary Methodologies
University of Warwick

September 2015

Table of Contents

Table of Figures

Acknowledgements

This thesis would have not been possible without the constant support and guidance offered by my two supervisors Celia Lury and Noortje Marres. I'm extremely grateful to Celia Lury for having showed me modes of doing social research which are expansive, inquisitive, rigorous and interdisciplinary and for a style of supervision which was at all times intellectually stimulating, rewarding and generative. I'm thankful to Noortje Marres for having decided to accompany and follow through this project in its different stages, and for her thoughtful, generous and challenging comments and observations that have largely enriched the conceptual and empirical development of my work throughout. To both, I am indebted for being a constant source of inspiration and for having me allowed to move with and learn from their academic trajectories throughout all these years.

I am thankful for the financial support provided by the Economic and Social Research Council (ESRC) that made possible to complete both my Masters and Doctoral research. My PhD initiated in 2010 in the Sociology Department at Goldsmiths College and has witnessed the founding of the Centre for Interdisciplinary Methodologies at the University of Warwick where it was completed. I am grateful to all members of staff and colleagues that I have met at both institutions who have contributed to the development of this thesis and who have expanded my doctoral training and thinking in interesting and unexpected ways. I am particularly grateful to Chiara Bernardi, Esteban Damiani, Tuur Driesser and Pablo Velasco Gonzales for stimulating conversations and exchanges and for the moral support they offered throughout the PhD. I am also indebted to Olga Goruniova for our illuminating conversations at Berkeley; to Emma Uprichard for a Skype session

which unearth the notion of data units in revealing ways; to Lisa Adkins for enabling a student research seminar atmosphere which enabled some of the concepts in the thesis to unfold and take shape; to David Stark for his timely and insightful comments on my writing style; and finally to Nate Tkacz and Will Davies for having allowed me to be part of an innovative project and a profoundly stimulating research environment during the last year.

I am also grateful to colleagues at other institutions who have invited and attended the presentation of my work at seminars and workshops, who have commented on early drafts and who have encouraged me to think and write about my research through different perspectives. Special thanks here go to Antonia Walford, Mark Carrigan and Jose Ossandon.

Personally, I would also like to thank my brothers Tomas and Enrique, because our unique and enduring relationship has been central in supporting me throughout these years in ways they might not be able to grasp completely from another continent. I'm also extremely indebted to my friend Lucia Ariza who has provided an invaluable amount of intellectual stimulation and warm understanding and whose sensibility has guided and enriched this research project in multiple ways. I am truly grateful to Daniel Chernilo too whose advice, patience and stake on British academic life have been extremely important to me at crucial times. I am also indebeted to Daniela Asikian and Maia Lemlij who have been supportive and understanding at times of doubt and struggle; to Diana Jaramillo, for her warmth and groundedness; to Daniela Jara, for her calmness and affectionate perspective throughout; to Paula Mena, for being a close confident of PhD and other matters of life; to Veronica Millenaar, for

all these years of friendship and academic entanglements at a distance; and to Olivia Kay, Averil Broad and Jane Guilmour for accompanying and encouraging me throughout. To all of them, I am grateful for having witnessed and helped me navigate the conflation of motherhood and doctoral research in highly creative and (sometimes) enjoyable ways.

Finally, this thesis would have never been possible without the constant support, encouragement and steadiness of my partner Rodrigo. I am grateful for Rodrigo's calmness and kindness at difficult times and for his relativist take on doctoral life and its traumas. Narrating the progression and completion of this thesis as a family effort and project throughout these years is entirely of his authorship and I am truly grateful for this.

This thesis is dedicated to our fantastic children Sofia and Felix. This is for and because of you two, and for all the love, happiness, fun and learning that you make possible.

Declaration

Some of the materials presented in Chapter Three are forthcoming in Gross, A. (2015), Search and the Ontological Capacities of the Digital, *Cultural Anthropology*, 30 (3). Parts of Chapter Three have been also published in Gross, A. (2015), The Domesticated Aboutness of Big Data Types, *Discover Society*, Issue 23.

Parts of Chapter Four have been published in Gross, A. (2012), The Economy of Social Data: Exploring Research Ethics as Device, *The Sociological Review*, 59: 113–129. The historical description of the development of the principle of informed consent present in Chapter 4, pages 136, 137 and 138 has been adapted from my Social Research Masters Degree dissertation.

Parts of Chapter Five and Chapter Six have been published in Lury C. and Gross. A (2014), The Ups and Downs of the Consumer Price Index in Argentina: From National Statistics to Big Data, *Partecipazione e Conflitto* 7(2): 258-277. This paper was collaboratively written and is equally attributable to both authors.

I hereby declare that this thesis is the result of my own work and has not been submitted for a degree at another university.

Abstract

This thesis contributes to the sociology and anthropology of data by examining the techniques and devices that are deployed to frame data as part of methodological, ethical, economic, digital, journalistic and artistic practices. The thesis starts by tracing the lineage of the concept of *frame* as part of the traditions of cybernetic anthropology, artificial intelligence, social interactionism and science and technology studies to delineate a conceptual framework that can account for the contextualisation of data. Empirically, the project focuses on two *data leaks* and repurposes the materials that emerged from these as case studies that render visible how different techniques and devices make possible the formation of two distinctive data types: *personal data* and *prices*. The first case study examines the making and unmaking of search keywords as personal and it is based on the materials that arose from the leak of a search engine database in 2006. This case study looks at how techniques like *reidentification demonstrations* and *data sequencing* have contributed to define search keywords as being about and capable of signalling persons while also investigating how ethical devices like *informed consent* and *anonymisation* work to depersonalise data instead. The second case study compares compositional against disaggregated framings of prices and it is based on the materials that became available as a consequence of the attempted disclosure of the databases used to estimate a national inflation indicator in Argentina since 2006. This case study explores how *product identification* and *data aggregation* techniques contribute to frame the fluctuation of prices as part of the measurement and communication of national statistics while also studying digital *scraping* and *imaging* as devices that frame the observation and interpretation of retail price variation for financial use.

Introduction

Data Contexts

In recent times both *data* and *databases* have emerged as relevant and crucial objects of study for the social sciences and the humanities. Since I started my research project back in 2010, a range of specialised books, journal articles, academic pamphlets and provocations, dedicated journals, and special issues have come to address the various forms and formats of data and databases[1] as objects of empirical investigation in their own right. These relatively recent academic advancements have emerged as part of a range of disciplines such as anthropology, sociology, history, media studies, science and technology studies and geography, amongst others. Such studies draw our attention to how databases and data are now so fully integrated into the dynamics of contemporary culture and society that they routinely become involved and participate, albeit often unnoticeably, in the formulation, enactment and definition of a wide range of objects, phenomena and experiences in and of the world. This body of work not only points to the historical constitution of data and databases, describing their technical specificities and accounting for their production. In also

[1] See for example Gitelman (2013); Kitchin (2014); Halpern (2015); Edwards (2010). Other earlier books that have directly or indirectly addressed data and databases as objects of study and commentary include Manovich (2001); Bowker (2005) ans Boker and Star (1999). With regards to articles see Mackenzie (2012); Mackenzie and McNally (2013); Ruppert (2012); Beer and Burrows (2013); Knox et al. (2012); Fuller and Goffey (2012b) and Amoore (2011) among others. For a pamphlet see Boellstorff and Maurer (2015) and for a provocation see boyd and Crawford (2012). The journal Big Data & Society was founded in 2014 and it is now one of the main outlets for research related to the social study of data. Computational Culture Issue 3 in 2013 was entirely dedicated to the study of databases.

studying the material and technical affordances as well as the representation, circulation and procedural operations of databases and data, this incipient but still somehow scattered research programme has also crucially foregrounded the agential, worlding capacities and effects of databases and data.

This thesis contributes to this emerging field of study by investigating how the relation between *data* and *contexts* becomes configured in and through the workings of different devices, techniques and methods. It is guided by the following question: *how does the articulation of contexts enable data to become about or related to particular entities and phenomena in the world?* As part of this emergent interdisciplinary field of study, the thesis makes clear that databases and data are objects that shape the composition and organisation of contemporary society and culture in significant and often problematic ways.

In 2006, a series of dramatic events unfolded after an America Online (AOL) database leaked, revealing approximately three million search queries undertaken by five hundred thousand web-search users over three months. In the same year, but approximately five thousand miles away in Argentina, a technical and legal dispute arose over the request made by government ministers to statisticians to disclose the database used to calculate inflation rates for a particular geographical district in the country. From search keywords spilling out from corporate research, to the potential breaching of databases containing information about the measurements of a public inflation indicator, this research project repurposes these instances to consider them as *empirical occasions* (Marres 2013). Data leaks are empirical occasions through which the contextualisation of data, and the devices, techniques and methods that

make such contextualisation possible, become visible and legible for social analysis. The thesis therefore examines, through two case studies, how two different data types - *personal data* on one hand and *prices* on the other - become contextualised through the actions of a series of devices, techniques and methods deployed by a variety of actors as part of a range of digital, ethical, economic, journalistic, methodological and artistic practices. It studies what these devices and techniques do, and how they contribute to make and unmake data and its contexts.

Before explaining the thesis and its structure in more detail, I would like to take two brief etymological detours. Etymologically, the word *data* can be traced to the Latin word *dare*, 'to give'. Daniel Rosenberg (2013) has historiographically studied the semantic use of the word *data*, demonstrating how the emergence and evolution of this word contributed to open up a conceptual space for the later development of twentieth-century innovations in information technology. According to Rosenberg, the term data was first used in the English language in the seventeenth century, when its meaning became closely associated with the development of scientific knowledge and expertise, which was evolving alongside new modes of producing, presenting and debating scientific knowledge (Kitchin 2014). In the late eighteenth century, Rosenberg (2013) notes however that the semantic function of the word *data* became decoupled from the use of the words *facts* and *evidence.* The meaning of *data* emerged instead as 'independent of any consideration of ontological truth' (2013, 18). A semantical distinction was thus historically produced between the words *data* on one hand, and *facts* and *evidence* on the other. This semantic distinction, Rosenberg suggests, might explain why 'when a fact is proven false, it ceases to be a fact' whereas 'false data can be data nonetheless' (Rosenberg 2013,18). Rosenberg

goes on to note that ‘facts are ontological, evidence is epistemological and data is rhetorical’ (2013, 18).

According to Rosenberg (2013), data came to exist semantically as prior to argument or interpretation; as that which is given and taken for granted but which can nevertheless become shaped in meaningful and partial ways by the deployment of rhetorical (and other types of) devices and techniques. This thesis similarly argues that the ontological status of data is heavily dependent upon the configuration of contexts. As Bruno Latour (1999) suggests in this respect, one should never therefore think of pre-given data, but of *sublata* instead - an expression that accounts for the material and semiotic achievements that bestow data with particular qualities and capacities, and the orchestration of the scenography that configures data as being about entities and phenomena in the world.

Like data, the word *context* has semantically evolved to denote that which is taken for granted as the background of different phenomena, events, interactions or objects. However, another brief etymological detour offers a different perspective on how contexts might be understood:

> The etymology of the word context suggests a derivation of the Latin word texere, ‘to weave’. The related Latin word *contextere* carries the meaning of ‘to weave together’, ‘to interweave’, ‘to join together’ or ‘to compose’ (*The Shorter Oxford English Dictionary*). Obsolete meanings of the word, therefore, range from ‘the weaving together of words and sentences, and literary composition’; to ‘the connection or coherence between parts of a

> discourse'. More contemporary meanings suggest 'the parts which immediately precede or follow a particular passage or text and determine its meaning'. (Dilley 1999, 4; italics in original)

Roy Dilley (1999) draws attention to how contemporary uses of the word context - including as an analytical strategy in many ethnographic studies and social scientific positivist accounts – have tended to move away from the original etymological meaning of the word, towards an understanding of context 'as self-evident, as a given attribute in the world, something that is stable, clear and sufficient and not requiring any qualification of its own' (1999, 2). Dilley (1999) proposes instead that contexts are mechanisms and processes that are deployed to make particular connections and disconnections. He argues that entities and phenomena become connected in particular ways to their surroundings, and that therefore contexts should be understood as the coordination of connections that are weaved together to foreground what ultimately becomes relevant and meaningful in the world. One could thus think of contexts not as predefined and fixed containers, but rather emerging as part of the relational, dynamic and continuously shifting properties that are seamed through activity and interaction.

This understanding of contexts as emergent from social activity can be traced back in sociology to the traditions of interactionism and ethnomethodology (Goffman 1974; Garfinkel 1967), which have advanced the conceptualisation of contexts not as already 'out there' waiting to be described or identified, but rather as situated and temporary achievements dependent upon human face-to-face interaction, communication and exchange. Whereas positivist accounts of context take entities

and activities as a priori, separate, and distinct from a given and already defined context, interactional and ethnomethodological accounts have pointed out that they are in fact woven together *to become seperate,* as a result of interactions. Moreover, thinking contextualisation as a phenomenon that is accomplished in interaction also entails considering contexts as being made *together with* the entities, activities and issues at stake.

Why is it then relevant to think *data* in relation to *context*? Reassessing what context is in the light of the current contemporary propagation of data, and rethinking data in the light of interactionist understandings of context, is important because contexts are becoming technologically articulated and mediated through data. On top of this, a myriad of old and new devices, technologies and methods are simultaneously redefining the contextualisation of data as an entity in itself. Technical contextual awareness and identification, for example, are now being made possible by the integration of big data, software, data analytics, sensors and ubiquitous and mobile computing (see Scobel and Israel 2013). The design and technical development of many of the current computational mechanisms and analytics being deployed to detect contextual signals as part of contemporary recommender systems are underpinned by the assumption that contexts are given and can be found 'out there' in the signalling of a myriad of data types (Seaver 2015). These systems strive to identify and apprehend contextual signals such as time and location (as well as ambient noise levels, heart rate spikes or quantity of emails unread in a given inbox), in order to provide more accurate and personalised recommendations. The assumption that these systems make is that by detecting what is assumed to be a *fixed context* into which a given user is inserted, then a more relevant recommendation of

consumption, decision-making, or life style choice becomes possible (Seaver 2015). This constitutes a representational and positivist mode of understanding context as a fixed and stable container for activity and entities, and data signals and their readings are thus seen as capable of pointing at the contours of such a container (Seaver 2015). Moreover, from this perspective, the methods, techniques and devices deployed to apprehend context are thought to be outside of it, rather than being part of its ongoing accomplishment and sustainment.

Thinking data in relation to context is important therefore because in many settings, data is being used as a signal with the capacity to express the context which 'contains' system users in space and time. This is of course only one of the multiple ways in which new technological formations accomplish context. However, data is not only being technologically shaped into contexts, but is itself also being contextualised as an entity which forms part of contemporary, everyday settings and practices. As the etymology of the word suggests, data is heavily dependent on what Kristin Asdal and Ingunn Moser (2012) describe as *practices of contexting*. What data is about and the relations that it establishes with a range of other entities and phenomena in the world is dependent upon the practices, methods, techniques, and material devices that contribute to accomplish, sustain or redefine its context. Although the accomplishment of context in interactionism and ethnomethodology was seen as the product of human face-to-face interaction and the outcome of linguistic interchanges, it is important to note here that this thesis also draws on the particular analytical sensibility of actor network theory, which takes into account how realities emerge through the agency and work of various instruments, materials, protocols, technologies, objects and bodies, human as well as non-human (Callon

1986). The thesis thus analytically engages with a series of devices, techniques and methods that aid in the contextualisation of data. It argues that by contextualising data, these devices, methods and techniques organise and shape the qualities and the material and semiotic capacities of data, contributing to define what data is about and how it becomes related, connected and disconnected to different entities and phenomena in the world.

The thesis also proposes that revealing the mechanisms by which data contexts are established matters because it makes visible both the *ontological politics* and the *political economy* of data. We can trace how the production of contexts enables data to become about and related to some entities and phenomena and not others, a partiality that can be understood in terms of what Annemarie Mol (1999) defines as *ontological politics*. Mol (1999) argues, based on the philosophical proposition that there are multiple possible realities, that the politics of ontological performativity reside in privileging the enactment and visibility of one reality over other possible enactments and versions. Studying what makes data contexts particular and partial, what delimits the relations that data establishes with other entities and phenomena, and the ways in which data becomes associated with particular qualities and capacities (and not others) is one way of accounting for the ontological politics of data. What is more, tracing the contextualisation of data as it occurs through the work of a range of methods, techniques and devices is also a way of making sense of the *political economy* of data; that is, of the configurations that enable data to become economically valuable, and the types of relations that need to be sustained or cut in order for such valuation to take place. Studying the contexting of data is therefore also a means of exploring its commoditisation, appropriation and ownership.

The thesis starts in Chapter One by conceptualising the notion of context. It does so by drawing on the notion of *frame* as developed in cybernetic anthropology, artificial intelligence, social interactionism and science and technology studies. I draw on these traditions, in particular the social interactionism of Erving Goffman (1974), to develop a conceptual framework that throughout the thesis illuminates how the configuration of framings contribute to shape the *aboutness* of data. Frames and framing devices enable data to become meaningful as being about particular entities and phenomena in the world, and not others. Frames and framing devices, I suggest, also coordinate particular modes of observing and visualising data, either collectively or individually. Frames establish connections and disconnections, and continuities and discontinuities, between data and other entities, and they enable totalising or partial perspectives of the composition of data and different phenomena. Finally, frames and framing devices are also considered throughout the thesis as mechanisms that attribute data with certain qualities and capacities.

Although the notion of frame as developed by Goffman (1974) is useful to think about the contextualisation of data, in Chapter One I also engage with the work of Michel Callon (1998), and other science and technology studies scholars, in order to consider the materiality and instability of frames. On one hand, I dwell on the notion of *device* (Callon et al. 2007) in order to take into account how materials, techniques and methods – and not only linguistic or interactional exchanges – participate in the framing of data. On the other hand, I also take into account the notion of *overflow* (Callon 1998), in order to develop a particular analytical sensibility towards the instability rather than the containment and durability of frames. Finally, Chapter One

also looks at three different accounts that have directly or indirectly addressed the contextualisation and decontextualisation of different data types: discourse (Baumann and Briggs 1990); opinions (Lezaun 2007); and soil sampling (Latour 1999).

Chapter Two presents the design of this research project. The Chapter starts by considering data leaks as a type of overflow. It then moves on to suggest how data leaks can be repurposed in social research if understood as empirical occasions (Marres 2013), through which the framing of data becomes visible, and the workings of framing devices more legible for social analysis. The Chapter discusses the aspects of the research design that led to the development of two case studies that allow me to analyse the framing of data. The Chapter argues that instead of taking empirical advantage of data leaks as society's experiments, it is methodologically more productive to *case* them as a way of studying data frames and the devices that enable them. The Chapter therefore considers what the epistemological and ontological implications are of understanding data leaks as a form of *societal experiment*. It then goes on to describe what was methodologically implicated in the design and configuration of two case studies based on the reorganisation of materials and devices that became visible as a consequence of the two data leaks I briefly described above.

The thesis subsequently presents the two case studies, each of which engages analytically with the devices, techniques and methods that frame the emergence of two distinct data types: personal data on one hand, and prices on the other. The first case study – entitled *The Making and Unmaking of Personal Data* – is concerned with how certain data units, like search keywords, are made personal or impersonal through the working of a range of devices and techniques. It is based empirically on

AOL's leaked database of search keywords, and what this leak made visible. This case study is comprised of Chapter Three – entitled *The Technical Redoing and the Aesthetic Reordering of Search Keywords* – and Chapter Four – entitled *Ethics Devices and the Economisation of Personal Data*. Chapter Three explores how search engines normally configure and contextualise the relation between search keywords and persons. However, it also studies how extraordinary public techniques like *reidentification demonstrations* - which compose persons out of anonymised databases - and *sequencing* - the reordering of database series in a sequential fashion - also contribute to frame search keywords as personal. This Chapter makes the political ontology of data visible by focusing on the frames and devices which fix data units as pertaining to particular entities like persons, and to phenomena like intimacy and interiority. Subsequently, in Chapter Four, the focus shifts to the political economy of personal data. In this Chapter, I dwell on two devices that also became visible as part of the search keyword data leak: *informed consent,* the type of decision taken by research subjects for participating in data extraction procedures; and *anonymisation*, the deletion of personal traits and traces in data. The Chapter suggests that ethics devices like these configure data in such a way so as to make it economically valuable.

The second case study, which spans Chapters Five and Six, is entitled *Price Formations and Statistical Compositions.* It looks into how the framing of price becomes methodologically, technologically, legally and visually accomplished by investigating some of the devices and techniques used to measure and communicate inflation. This case study is based on the attempted data leak that took place when government officials requested statisticians to disclose the database used for

calculating the national inflation indicator in Argentina in 2006. Here, I focus specifically on a series of devices and techniques that became visible as a result of this attempted leak. Chapter Five – entitled *The Methodological Formation and the Secret Containment of Price* - analyses two devices that operate in tandem to articulate the framing of prices as *pure,* for the purposes of measurement and communication of an inflation indicator. Firstly, the Chapter looks at one of the methods deployed as part of price collection mechanisms, called *product identification*. It then moves on to analyse the *aggregation* of statistical data as a communication device. As in the first case study, this Chapter is more concerned with the ontological politics of data by showing how the *aboutness* of price depends on its framing. In Chapter Six - entitled *Digital Methods and the Financialisation of Retail Price* - the analysis again seeks to shed light on the political economy of data. It looks at two alternative commercial inflation indicators that emerged directly or indirectly as a consequence of the controversy generated in Argentina over the national inflation indicator. The Chapter suggests that the digital devices which frame prices as part of these commercial initiatives reframe the aboutness of price in economically valuable ways. Firstly, the Chapter looks at *scraping*, a semi-automated digital technique for the collection and annotation of online prices. Secondly, it analyses *imaging*, that is, the annotation of prices that measure inflation via digitally crowd-sourced and dispersed images. In accounting for the framing of prices as part of private initiatives, the Chapter looks at the changing function of official economic indicators and the social challenges that alternative indicators of inflation poise.

In the conclusion, I reflect upon the relations between frames and data types and functions. I suggest that the thesis either implicitly or explicitly points at how the

function of traditional data types like *personal names* and *official statistical indicators* - understood as coordinators of social and economic action - might be changing. I point at two recent empirical occasions that can serve to further expand in the future the incipient conceptual proposition developed in the thesis: that certain data types have become established to effect social coordination functions and that these functions are being challenged by the emergence of new technological arrangements, data types, data frames and analytical techniques among others.

Chapter One

The Framing of Data

Introduction

My thesis investigates the *framing of data*. It presents and explores some of the methods, techniques and devices that are being deployed to frame data - that is, to attribute data with contexts and make data about different entities and phenomena in the world. The thesis starts from the assumption that *data* is an ontologically unstable entity, and that *frames* – and also importantly the devices, methods and techniques that contribute to define them – work to delineate and organise the *framing* of data. In simple terms, the thesis is based on the proposition that what data is about and related to is the outcome of framings and *practices of contextualization* (Asdal and Moser 2012). In other words, thinking data alongside framing raises the possibility that data becomes formed, and its ontology fixed, through the establishment of relations that extend beyond itself.

In the following Chapter, I describe and engage with the theoretical framework that underpins the thesis in order to conceptualise what is that enables data to alternate and *become about* different phenomena and things in the world. The theoretical framework here presented draws on the notion of *frame* as developed in the traditions of cybernetic anthropology (Bateson 1955, 1972); cognitive frame theory (Minksy 1975); social interactionism (Goffman 1974); and the later adoption and

reformulation of the term in science and technology studies (STS) and actor-network theory (ANT) (Callon 1998).

The first section of the Chapter looks at the lineage of the concept of *frame* and its development in the traditions of cybernetic anthropology, cognitive frame theory and social interactionism. It argues that the emergence and use of the term *frame* in these fields poses a challenge to positivist understandings of what contexts are and do. This first section of the Chapter then goes on to expand the conceptualisation that Erving Goffman (1974) developed for understanding frames as mechanisms for the *production of ontological aboutness*. It traces how Goffman (1974) relied on Gregory Bateson (1955, 1972), William James (1869) and Alfred Schutz (1945) to develop his own particular conceptual proposition about frames, understood as interactive organisations of experience that serve to define both the meaning of given situations and actors' awareness of it.

The Chapter then moves to focus on how the term *frame* has been adopted in STS, particularly in the work of Michel Callon (1998). It looks at the reformulation of the term by Callon (1998) through his elaboration of two other closely related concepts: that of *overflow* and that of *device*. Whereas Goffman (1974) focused on studying the types of containments enabled by frames, Callon (1998) develops a particular analytical sensibility towards their instability, as well as pointing to the different associations that can always be established between that which is considered to be inside and outside of frames. Additionally, the elaboration of the term *device* within this tradition is highly useful in incorporating materials and objects – and their agency - into the study of framings.

A third and final section of the Chapter engages with the development of the concepts of *inscriptions* and *immutable mobiles* as part of the work of Bruno Latour (1999); the description of processes of entexting and contexting discourse in Richard Bauman and Charles Briggs (1990); and the framing of opinions in the work of Javier Lezaun (2007). This section of the Chapter pays particular attention to the processes of contextualisation and decontextualisation that have been described in these accounts as fundamental for the ontological stability, production and mobility of different types of data.

Frames

In this section of the Chapter, I engage with conceptualisations of *contexts* understood in relation to *frames*. Let me start by tracing the concept of *frame* as developed by Bateson (1955, 1972). In *Steps to an Ecology of Mind*, Bateson (1972) elaborated on different forms of cognition, learning and psychological pathologies that emerge as the result of environmental interactions and communication patterns. Bateson (1972) offered a bio-cybernetic proposition of the mind, not as contained in the brain, but as emergent from and configured by the complex relationship established between the perceiver, its communicative interactions and the environment (Parisi 2012). His proposition was radical at the time as it discarded the idea of a mind contained within an individual psychology be, and replaced it with 'an ecology of ideas in systems or "minds" whose boundaries no longer coincide with the skins of the participant individuals' (1972, 339). This theoretical (but also

methodological)[2] reformulation of minds as (partly) external to bodies required a further re-elaboration of the notion of contexts or environments, and their functions and relations, and this is the reason why a theory of mind might be relevant to the study of data contexts. In this respect, Bateson (1972) offered some interesting conceptualisations of the notion of context(s),[3] but it is in particular the notion of *frame* conceived of as a mechanism for active contextualisation and bounding that is most relevant, and that I am here interested in expanding further.

In his essay *A Theory of Game and Fantasy* (1955), Bateson cut across different disciplines - from mathematics, philosophy, communication theory and psychology - to explore the concept of the *frame*. Bateson (1955) described frames as the exchange of both verbal and non-verbal cues among humans and animals that indicate how ongoing interaction should be understood, and that configure what interaction is ultimately about. He suggested that individuals exchange signals that allow them to agree upon the level of abstraction that messages are intended to be interpreted within. In other words, he proposed that individuals (but also animals) produce meta-communicative signals about the signs they happen to exchange as part of a particular given interaction, and these meta-communicative signals *provide a context* for the exchange of signs to occur. Bateson (1972) later suggested that understanding an action or utterance as occurring 'in' a context should be conceived as an heuristic error. He noted that this conventional way of talking about things or events as if happening 'in' a context is dependent upon understanding a particular thing as a

[2] See Halpern (2012).

[3] See for example the development of experimental contexts of learning typologies advanced in Bateson (1972).

‘dependent variable’, whilst its context is conceived of as an ‘independent variable’ (339). Bateson argued that this particular view of how a thing is related to its context does not position the particular thing ‘as part of the ecological subsystem called context’ but rather ‘as the product or effect of what remains of the context after the piece which we want to explain has been cut out from it’ (1972, 338).

Bateson invoked two analogies (1972) as entry points to think about frames. The first abstract analogy was that of the mathematical set. He noted with regards to this that

> In set theory mathematicians have developed axioms and theorems to discuss with rigour the logical implications of membership in overlapping categories or “sets”. The relationships between sets are commonly illustrated by diagrams in which the items of or members of a larger universe are represented by dots, and the smaller sets are delimited by imaginary lines enclosing the members of each set. Such diagrams thus illustrate a topological approach to the logic of classification. (1972, 186)

It followed from this, for example, that the ‘play’ of two individuals on a given occasion could be defined as the set of all the messages exchanged by them; and that in a set-theoretical diagram these messages could be represented by dots and enclosed by a line to distinguish and separate them from non-play messages (1972, 186). In these terms, frames were conceived as modes of ‘referring by ordering’, in the sense that ‘a frame always sorts things as either belonging or not belonging and this process is mediated by axioms or principles – indeed the axioms are what define the frame; they are the conditions of its possibility’ (Tackz 2012, 69).

Bateson (1972) also used the analogy of the picture frame to consider a different set of issues. He noted that

> The frame around a picture, if we consider this frame as a message intended to order or organize the perception of the viewer, says, "Attend to what is within and do not attend to what is outside." Figure and ground, as these terms are used by gestalt psychologists are not symmetrically related as are the set and non-set of set theory. Perception of the ground must be positively inhibited and perception of the figure (in this case the picture) must be positively enhanced. (1972, 187)

Bateson continued:

> The picture frame tells the viewer that he is not to use the same sort of thinking in interpreting the picture that he might use in interpreting the wallpaper outside the frame. Or, in terms of the analogy from set theory, the messages enclosed within the imaginary line are defined as members of a class by virtue of their sharing common premises or mutual relevance. The frame itself thus becomes part of the premise system. Either, as in the case of the play frame, the frame is involved in the evaluation of the messages it contains, or the frame merely assists the mind in understanding the contained messages and reminding the thinker that these messages are mutually relevant and the messages outside the frame may be ignored. (1972, 188)

The analogy of the picture frame demonstrates that frames do not only include and exclude elements and are therefore generative of classes and categories of things, phenomena or occasions but also that frames distribute and organise perceptual and interpretative understandings of what is inside and outside of them. For Bateson (1972), frames informed what might become *perceptually* visible as something coherent as a unit and distinguishable from a background or context. In producing this difference, frames enact the inside and the outside, the figure and the ground, as perceptually distinguishable and separate.

Whereas Bateson (1972) considered frames as primarily communicative, constitutive of interaction and articulated outside of individual cognitive structures, alternative developments in the field of cognitive theory conceptualised frames instead as mental structures or schemas ingrained in human cognition. *A Framework for Representing Knowledge* - an essay authored by Marvin Minsky and first published as part of an edited collection on the *Psychology of Computer Vision* in 1975 – could be considered as one of the seminal theoretical contributions of what was later defined as cognitive frame theory (Dewulf et al. 2009). Formulated within the field of artificial intelligence, the tradition advanced by Minsky (1975) focused on investigating 'cognitive frames as mental structures that facilitate organizing and interpreting incoming perceptual information by fitting it into already learned schemas or frames about reality' (Dewulf et al. 2009, 158). In the same article, Minsky claimed that frames are remembered frameworks or structures that are selected when encountering new situations, or adapted to fit reality by changing details as necessary. He proposed that

> A frame is a data-structure for representing a stereotyped situation, like being in a certain kind of living room or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed. (…) We can think of a frame as a network of nodes and relations. The "top levels" of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have many *terminals* -"slots" that must be filled with specific instances or data. Each terminal can specify conditions its assignments must meet. (…) Collections of related frames are linked together into frame-systems. The effects of important actions are mirrored by transformations between the frames of a system. These are used to make certain kinds of calculations economical, to represent changes of emphasis and attention, and to account for the effectiveness of "imagery"'. (Minsky 1975, 33, italics in original)

Aligned to his work on artificial intelligence, Minsky's (1975) objective in uncovering the workings of the human cognitive frames used for understanding and navigating the world was to develop machines that would display human-like abilities. Minsky (1975) proposed that learnt, stored and already organised information and knowledge about the world is cognitively put to the test whenever an individual encounters different arrangements, be these spatial, social or linguistic. For example, the frame of a given living room includes certain information, such as that the room has walls, a ceiling, lights and a door. The door can be thought of as a slot that accepts values such as 'wood' or 'metal', but it does not accept values such as 'liquid'. Therefore, if a human or even an artificially intelligent machine enters a

particular living room, the learnt or structural frame –the generic cognitive frame - of what constitutes a living room is applied and put into negotiation with the specificities and particularities of individual living rooms and their possible multiple slots.

Whereas cognitive frame theory conceived of frames as representations stored in memory - and 'framing' as a process in which certain cognitive data structures are applied or tested against particular situations – a third tradition[4] very closely aligned with Bateson's conceptualisation considered frames as the outcome of the dynamic shaping of meaning and interpretation in ongoing social interactions. The work of Erving Goffman entitled *Frame Analysis: An Essay on the Organization of Experience* (1974) can be situated in this later social interactionist tradition, and his work on frames is one of the most thorough and expansive attempts to advance the sociological conceptualisation of contexts. Arguably, the earlier work of ethnomethodologist Harold Garfinkel (1967) had already proposed that contexts should not be accepted as sociological explanatory tools but rather analysed as enactments emerging in the unfolding of interactions. From such an ethnomethodological perspective, contexts are seen as held, mobilised, recreated and challenged by actors in order to make sense of their actions, and are not conceived of as phenomena outside of, or external to, a given interactional activity taking place in a particular space and at a particular time. Therefore, in both ethnomethodological

[4] For a description and comparison of the divergent ontological, epistemological and methodological assumptions carried forward in both traditions see for example Dewulf et al. (2009).

and interactionist accounts, empirical and conceptual interest focused on how contexts were *accomplished*.

The move towards understanding context as an accomplishment dealt with two inherent problems of positivist conceptualisations of contexts identified by Bauman and Briggs (1990): that of inclusiveness, and that of false objectivity. Bauman and Briggs (1990) argued that traditional definitions of contexts, which propose that everything lying outside a given object, event or interaction is potentially its context, are overly inclusive and therefore imprecise about the point at which an adequate range of contextual factors has been achieved - or in other words, where it is that context stops.[5] Additionally, Bauman and Briggs also noted the problem of false objectivity that emerged from most of the positivistic definitions of contexts that they analysed. They pointed out, with regards to linguistic utterances for example, that

> Positivistic definitions construe context as a set of discourse-external conditions that exist prior to and independently of the performance. This undermines the analyst's ability to discern how the participants themselves determine which aspects of the ongoing social interaction are relevant. It also obscures the manner in which speech shapes the setting, often transforming social relations. Reifying "the context" also implicitly preserves the premise that meaning essentially springs from context-free propositional content, which is then modified or clarified by "the context". (1990, 68)

[5] Bauman and Briggs argued that since it would be impossible to take into account all aspects of the context, 'the researcher becomes the judge of what merits inclusion' (1990, 68).

We can conclude that what Bateson (1955, 1972) and Goffman (1974) proposed in their accounts was a shift from a positivist understanding of context to a conceptualisation and description of the processes, devices and dynamics of contextualisation. In both accounts, contexts were not taken as the already-given physical and social environment in which particular interactions take place, but as emergent from negotiations between different agents partaking in social or communicational interaction. Contextualisation was conceived by Bateson (1955, 1972) and Goffman (1974) as an ongoing process which could be analytically discerned by 'attending to the "contextualisation cues" that signal which features of the settings are used by interactants in producing interpretative frameworks' (Briggs and Bauman 1990, 68).

Aboutness

In this section of the Chapter, I would like to particularly focus on Goffman's account of frames. Goffman (1974) conceptualised a frame as a principle of organisation that defines situations and guides experience in everyday life[6]. For Goffman (1974), frames cut strips out of the stream of ongoing everyday activity and render them intelligible. Frames demarcate events and the meaning they become attributed with.

[6]Goffman (1974) acknowledged the coinage of the term 'frame' to Bateson (1955). Goffman followed Bateson in assuming that 'definitions of a situation are built up in accordance with principles of organization which govern events – at least social ones – and our subjective involvement in them; frame is the word I use to refer to such of these basic elements as I am able to identify. That is my definition of frame. My phrase "frame analysis" is a slogan to refer to the examination in these terms of the organization of experience' (1974, 11).

In other words, frames can be understood as mechanisms for the production of aboutness.[7] Because they enact cuts and boundaries, frames and framing devices also establish everyday situations or events as having internal and external aspects, as both containing and contained (Strathern 2002).

I would like here to dwell on the notion that frames can be conceived of as mechanisms for the production of aboutness, and why such production might also be deemed ontological. Goffman (1974) understood frames as mechanisms for 'defining situations'. This, as for Bateson (1972), presupposed an understanding of cognition as happening outside individuals. Individuals were not thought of as independently generating the definition or meaning of a given situation, but were instead thought of as being provoked by the frame to produce a correct assessment and awareness of what the situation ought to be for them. Goffman claimed: 'I assume that when individuals attend to any current situation, they face the question: "What is it that's going on here?"' (1974, 8). Goffman (1972) referred to James (1869) in proposing to move away from questions of what reality is, to question instead *under what circumstances we think things are real*. He argued that

[7]The capacity to render aboutness has also been attributed to contexts in other ethnomethodological and social scientific accounts. For Garfinkel (1967) any action is doubly contextual in the sense of being context-shaped and context-renewing. Actions are context-shaped in that their 'contribution to an ongoing sequence of actions cannot adequately be understood except by reference to the context […] in which it participates' (Heritage 1984, 242). On the other hand, actions are context-renewing because they form the immediate context of the subsequent actions that they trigger, and therefore contribute to shape the frame through which future actions will be attributed with particular meaning.

> The important thing about reality he [James] implied, is our sense of its realness in contrast to our feeling that some things lack this quality. One can then ask under what conditions such a feeling is generated, and this question speaks to a small, manageable problem having to do with the camera and not what it is the camera takes pictures of (1974, 2).

According to Goffman (1974), Schutz (1954) also insisted that emphasis need to be placed on understanding and uncovering the conditions that need to be fulfilled so that particular realms of reality, or as Schutz put it, 'finite providence of meanings', are privileged over others (Schutz, cited in Goffman 1974, 3-4). He also suggested that attention should be paid to the rules which, when followed and sustained, generate a reality of a given kind (1974, 5). Goffman (1974) was particularly interested in uncovering the mechanisms that *define* situations as being real, but also particular things or phenomena. He paid close attention to the rules and devices that enable individuals to tell, agree, but also sometimes manipulate what a given situation or event is about. Frames thus establish a sense of aboutness: they organise cognitive and bodily[8] experience, and elicit an awareness of the meaning of a given situation, or as Goffman also described it, a strip of activity. In a rather Batesonian manner, Goffman (1974) went further to specifically claim that the organisation of experience is 'something that an individual can *take into* his mind' (1974, 13; my emphasis).

Let me here expand Goffman's conceptual framework further. In *Frame Analysis,* Goffman (1974) started from the assumption that taken for granted 'primary

[8] See again Goffman's take on Schutz (1945) for understanding the relation between bodies and frames (1974).

frameworks'[9] operate as part of everyday life. These primary frameworks are nothing more than frames which have become naturalised or ingrained as mechanisms for applying repeated boundaries to a world that would otherwise be experienced as amorphous and meaningless. Goffman claimed that a primary framework 'allows its user to locate, perceive, identify, and label a seemingly infinite number of concrete occurrences defined in its terms' (1974, 21). In line with other social scientific traditions, his work gained analytical purchase when considering a variety of situations where primary frameworks were made visible because expectations about them became unhinged.[10] I shall return to this point and elaborate on its implications from a methodological perspective later in Chapter Two.

What is important to note here however is that Goffman (1974) developed an analytical framework designed to identify the 'lamination of frames', or the

---

[9] One of the main criticisms made of Goffman's work has been of his failure to develop the notion of 'primary frameworks' as the 'framework base' in which subsequent frame transformations are produced (Manning 1992, 129-132). Goffman referred to primary frameworks as follows: 'when an individual in our Western society recognizes a particular event, he tends, whatever else he does, to imply in this response (and in effect employ) one or more frameworks or schemata of interpretation of a kind that can be called primary. I say primary because application of such framework of perspective is seen by those who apply it as not depending on or harking back to some prior or "original" interpretation; indeed a primary framework is one that is seen as rendering what would otherwise be a meaningless aspect of the scene into something that is meaningful' (1974, 21). This definition resonates with the cognitive frame theory tradition described above, as it refers to knowledge or interpretative schemas which are reproduced and ingrained as part of individual and collective memories but conceived of as 'in the mind'.

[10] He described for example 'astounding complex' or occurrences, the exhibition of stunts; fortuitousness; jokes and tense situations as instances or occasions which challenge, destabilise and make visible primary frameworks.

superimposition and layering of frames. In other words, Goffman's account was based on the assumption that primary frameworks can be *rekeyed;* that is, their patterns can be used and transformed in order to redefine their original meaning and aboutness. As Goffman argued, 'given the possibility of a frame that incorporates rekeyings, it becomes convenient to think of each transformation as adding a layer or lamination to the activity' (1974, 82). These layers or laminations ought to be understood not as properties or fixed structures imposed onto events or situations, but rather as multiple, relative and unstable positions brought into being by frames that are actively sustained as different to each other. Goffman (1974) argued that an activity perceived as organised in terms of a primary framework is subject to two basic types of interventions and vulnerabilities, which he went on to define as as *keyings* and *fabrications*.

Primary frameworks are *keyed* whenever 'their meanings are transformed into something patterned on, but independent of them' (Manning 1992, 123). Goffman (1974) distinguished and typified five basic keying mechanisms that usually laminate and therefore complicate primary frameworks in Western societies: make-believe

activities or games; contests; ceremonials; technical redoings; and regroundings.[11] For Goffman keyings might be subject to rekeyings and therefore there is no 'obvious limit (…) to the number of rekeyings to which a particular strip of activity can be subject to' (1974, 80).

In addition to keyings, primary frameworks can become altered by *fabrications*. A frame is fabricated when 'it is organised by a party in such a way that others actors will have false ideas of what is happening in the frame' (Manning 1992, 126). The differences between fabrications and keyings is that whereas in the latter actors interpret the frame in the same way because information about the framing process is evenly distributed and made available, in the former some actors are misled about what it is that is going on inside the frame either by the concealment of information by other actors or by the blurring of the rim of the frame altogether. In particular, the concealing, revealing and asymmetrical distribution of information define fabricated situations.

[11] An example of a primary framework rekeyed as a technical redoing for Goffman would be a demonstration. Goffman defined these situations as 'performances of a task-like activity out of its usual functional context in order to allow someone who is not the performer to obtain a close picture of the doing of the activity. This is what happens when a salesman shows how a vacuum cleaner works to pick up the dirt he has instructively dropped on a housewife's floor, or when a visiting public health nurse shows an unwashed mother how to wash a baby, or when field commanders are shown what a piece of artillery will do (…)' (1974, 67). Manning (1992) argued that technical redoing serves to simulate, in the here and now, the unfolding of a situation that is expected to happen in the future 'while accompanied by the instruction to treat the frame as if it were what it purports to be' (1992, 125).

In Goffman's (1974) terms, multiple frames and different degrees of laminations operate at any one time for any given activity occurring in any place. It follows that it is possible to address or identify at least two or more features of a given activity or interaction: on one hand, there is the innermost layering 'wherein dramatic activity can be at play to engross the participant' and on the other hand there is what he defines as the 'outermost lamination', or the rim of the frame, which serves to define 'what sort of status in the real world the activity has, whatever the complexity of the inner laminations' (1974, 82). Under such conceptualisation, frames could be seen as organising not only experience, awareness and meaning - but also as effective means of engaging and captivating actors' focus of attention and perspectives.

Overflows and Devices

Michel Callon (1998) is one of many[12] STS scholars who have found inspiration in the work of social interactionists and ethnomethodologists. Callon's work has engaged with the process of framing in and of the economy and economics by explicitly adapting the term 'frame' from Goffman (1974). Callon thus explains with regards to frames and the economy:

> Framing is an operation used to define agents (an individual person or a group of persons) who are clearly distinct and dissociated from one another. It also allows for the definition of objects, goods and merchandise which are perfectly

[12] See for example Mol (2003); Latour (2007); Suchman (1987) and more recently Marres (2012).

> identifiable and can be separated not only from other goods, but also from the actors involved, for example in their conception, production, circulation and use. It is owing to this framing that the market can exist and that distinct agents and distinct goods can be brought into play. Without this framing the states of the world cannot be described and listed, and consequently, the effects of the different conceivable actions cannot be anticipated. (Callon 1998, 17)

The term *frame,* in the work of Callon (1998), has a distinct advantage over the term 'limit' as it enables the separation of the inside and outside of phenomena - such as the inside and outside of that which is 'economic', for example - to be seen less as a fixed and given boundary and more as a process of 'partitioning, bracketing or reducing' (Lury 2004, 154). Callon (1998) in particular focuses on elucidating the relations and dynamics that can be established between out-of-frame and in-frame activity.[13] Callon's reading of Goffman (1974) takes into account that the process of framing has a dual nature, in that the frame 'puts the outside world in brackets, as it were, but does not actually abolish all links with it' (1998, 249). Callon sustains that

[13] Goffman (1974) also addressed inflows and outflows when he described instances in which frames are broken which he defined as moments of 'flooding' in and out of frames. He wrote, 'given that a frame applied to an activity is expected to enable us to come to terms with all the events in that activity (informing and regulating many of them), it is understandable that the unmanageable might occur, an occurrence which cannot be effectively ignored and to which the frame cannot be applied, with resulting bewilderment and chagrin on the part of the participants. In brief, a break can occur in the applicability of the frame, a break in its governance' (1974, 347). One could consider overflows as rekeyings as well, or as instances in which fabrications are made visible to all actors involved in a given framing process. Goffman sees frames as iterative and superimposed; in other words, there are no events or situations in Goffman's terms which are not ultimately part of a frame within other frames (or rekeyings).

far from creating a complete separation from an outside, framing would be inexplicable if there were not a network of connections with the outside world, or in Callon's words, 'everything mobilised in the framed setting guarantees, simply by virtue of its presence, that the outside world is also present' (1998, 250).

Callon proposes that there are two ways of analytically dealing with this double nature of frames: one is to take frames as the norm, and overflows (or inflows and outflows) as sporadic occasions - this is a view[14] that puts an emphasis on the closure of situations and the containment of overflows. The second way focuses on the 'omnipresence of connections with the outside world and the irrepressible and productive overflows that frames encourage' (Callon 1998, 250). He defines this stance as one in which overflows are the norm because framings happen to be arduous to sustain, are expensive, and are always imperfect. So while Goffman (1974) appeared analytically inclined towards the first view - that is, by focusing on boundaries and the way interactions tend to favour and sustain containments - Callon (1998) argues on the contrary that, as part of framing processes, overflows are not only common but also functional. He argues further that a complete disconnection of an entity or action in a given frame is not only practically impossible but also difficult

[14]Namely, that of economic theory, the setting of markets, and the containment or internalisation of externalities. Callon argued that 'framing defines the effectiveness of the market because, in this closed interactional space, each individual can take into account the viewpoint of every other individual when reaching a decision. In this sense, it is possible to assert that externalities are simply the result of imperfections or failures in the framing process. Yet in certain cases framing is either impossible to achieve or deliberately transgressed by actors: this produces overflows which cause the barriers to become permeable. Economic theory regards overflows as accidental and consequently that the framing should be perceived as the norm towards which everything should tend.' (1998, 251)

to orchestrate and sustain over time. Therefore overflows, subsequent frame laminations, juxtapositions and intersections are always active and latently waiting to happen.

For Callon, framings constitute 'violent efforts' to extricate agents from a variable network of connections, in order to push them into a fixed and 'demarcated stage' (1998, 253). Callon further claims that although this stage might be a temporary achievement, the links that are sustained with the outside world betray the stage's existence 'simply by the fact that the agents are simultaneously involved in other worlds from which they cannot be fully detached' (1998, 253). Callon uses the example of 'contracts' as a way of illustrating the double nature of frames. He claims

> No contract is capable of, or has an interest in, systematically suppressing all connections, burning all bridges or eliminating the dual nature of every element involved. Which is why the heterogeneous elements, that are linked together in order to frame the contract and its performance, in reality take part in its overflowing: and it is precisely because they are sources of overflow that they make the contract productive. (1998, 255)

Different elements that are included as part of the frame represent openings onto other frames to which they give access. There are no static, enclosed insides to frames, but only configurations that make possible the fleeting stabilisation of certain connections, that is, they contain elements that are later productive of destabilisations and spillages. Callon (1998) went on to suggest that these spillages or overflows, or

what are defined as ‘externalities’ in the context of economic theory, can ultimately become internalised as part of frames.

Callon (1998) also suggests that framing is a process of disentanglement. In particular with market transactions, he noted that the status of goods can change and that in order to transform any given object into a commodity, for example, it is necessary to ‘cut the ties between the thing and the other objects or human beings one by one. It must be decontextualised, dissociated and detached.’ (1998, 19). The notions of entanglements and disentanglements imply that specific investments and precise actions are required to cut certain ties and to internalise others. However, entanglements and disentanglements need to be understood as two sides of the same coin: disentangling entities from a given enacted context entails simultaneously investing entities with a newly performed surrounding. In other words, entanglements and disentanglements can be understood as processes of keyings and rekeyings, as conceptualised by Goffman (1974), and hence overflows could also be understood as a type of recursive rekeying.

Callon (1998) also noted that frames should not only be conceived as merely linguistic, communicative or human-interactional, but that they should also be thought of as material. When referring to one of the most common examples of frames and framing devices in Goffman’s work, that of the theatrical frame, Callon noted that ‘a whole series of material means are used to demarcate the theatrical space and the actions that take place within it: the building itself; the internal architecture; the bell; dimmed lights and raising of the curtain that indicates the start of the performance’ (1998, 249). As noted above, Callon (1998) and the strand of

economic sociology that is rooted in ANT more generally (see for example Callon et al. 2007; Mackenzie et al. 2007) extend this notion of material framing in order to understand the configuration of markets and how the devices and techniques that frame and provoke valuation and calculation (among other economic phenomena), emerge.

I would like briefly to engage further here with the conceptual development, in ANT and STS, of the notion of the *device* as a way of foregrounding the participation of objects and materials, both as part of frames, and as framing devices. Whereas Bateson (1955, 1972) saw meta-communicational signals as one of the mechanisms through which a frame became accomplished as part of ongoing communicational interactions, what is distinctive about STS approaches is an emphasis on the materiality of frames and framings. Callon et al. (2007) propose and elaborate on the notion of *device* in order to consider how objects, materials and technical means could be brought into the analytical frames of sociological enquiry. Callon et al. (2007) claim for example that devices could be thought of as objects with agency: 'whether they must just help (in a minimalist, instrumental version) or force (in a maximalist, determinist version) devices do things. They articulate actions; they act or they make others act' (2007, 2). This tradition has in particular investigated how objects, devices and materials frame economic action; how they provoke cognitive dispositions such as calculation; and also how they contribute to organise markets and agents in particular ways. Attending to the materialist sensibility of STS and ANT is important because these traditions have advanced the inclusion of non-human agents as part of framing processes.

Data Contexts (Revisited)

In this final section of the Chapter, I would like to point to some accounts that have specifically addressed processes of data contextualisation and recontextualisation. These accounts have looked into the mechanisms and processes that enable data to be produced and detached from particular contexts; and to circulate and become recontextualised as part of other settings and be deployed for a range of purposes. These accounts have also looked at the type of relations that data sustains with different settings and entities, and the mechanisms and the devices that work to frame data and stabilise its aboutness. Here I engage with accounts of how the detachment of textual and verbal discourse from situated interactions is achieved (Bauman and Briggs 1990); conceptualisations of how the formation of opinions is dependent upon the staging of focus group frames (Lezaun 2007); and empirical renderings that describe the processes and practices which enable scientific fieldwork data to circulate and become recontextualised (Latour 1999). By drawing on these conceptualisations and empirical undertakings, I intend to draw attention to accounts that have addressed the importance of contexts for the articulation and ontological stabilisation of data.

Firstly, Bauman and Briggs (1990) explore the entextualisation of textual discourse from situational contexts. They argue that while there are processes that anchor discourse in contexts of use, other mechanisms potentiate and enable its detachability. They account for how the transformation of interactional speech acts into material strings of text that are analysable as discourse take place. They define this process as

one of 'entextualisation' (1990, 73). Such transformation from situated discourse to material text was described as

> The process of rendering discourse extractable, of making a stretch of linguistic production into a unit, a text, that can be lifted out of its interactional setting. A text, then, from this vantage point, is discourse rendered decontextualizable. Entextualization may well incorporate aspects of context, such that the resultant text carries elements of its history of use within it. (Briggs and Bauman 1990, 73)

For Bauman and Briggs (1990), a succession of recenterings or recontextualisations might happen as part of a single event and can be either simultaneous or serial. They suggest that in mapping the dynamics and dimensions of their transformations, it is important to note whether texts are linked to previous renderings as either repetitions or quotations, or whether they are being deployed as performances, or as rehearsals.

As part of the STS tradition that pays particular attention to devices, Javier Lezaun (2007) describes some of the techniques that rendered the opinions elicited from consumer focus groups trade-able and detachable from these particular contexts. He claims that focus group opinions are 'statements that are generated in an experimental setting but that can be disseminated beyond their site of production' and that 'producing opinions of such value and mobility is a highly complex technical process' (2007, 131). Lezaun (2007) claims that the focus group frame and the devices that contribute to its imposition is what enables the elicitation and emergence of detachable and hence tradable opinions. He describes some of these devices: the

techniques deployed by moderators to balance and merge the apparently contradictory dimensions of naturalness and artificiality in the opinions that arise from the focus group experimental setting; how the focus group is designed and the moderators trained to generate an 'isegoric' situation (Lezaun 2007, 140), that is, a situation that generates non-rhetorical formulations of opinions; and how the material set-up of focus groups, including the one-way mirrors these are usually equipped with, works as a reflexive device for moderators to asses the real-time value of the opinions being prompted. Rather than treating opinions as mere expressions of beliefs, Lezaun points at how in focus group settings they are 'generated in a highly mediated fashion and through complex technologies of investigation', where the articulation of a context through the work of the moderators, and the materiality of the focus group setting, guarantees the extraction of opinions deemed valuable (2007, 147).

Arguably, these are processes that Bruno Latour (1999) was also inspired by when he develops the notions of inscriptions and immutable mobiles. He defines 'inscription' as

> A general term that refers to all the types of transformations through which an entity becomes materialized into a sign, an archive, a document, a piece of paper, a trace. Usually but not always inscriptions are two-dimensional, superimposable, and combinable. They are always mobile, that is they allow new translations and articulations while keeping some types of relations intact. Hence they are also called "immutable mobiles", a term that focuses on the

> movement of displacement and the contradictory requirements of the task. (1999, 306-307).

In his ethnographic account of the scientific sampling of soil in the Amazon forest, Latour claims that one should never speak of 'data' or what is given, but of 'sublata' or 'achievements' (1999, 42). Latour traces 'the achievements of immutability' that enable scientific inscriptions to be mobile and circulate without their epistemic and ontological status being altered by their means of transportation or their potential entanglement in different settings (1999).[15] Latour (1999) was also concerned with how inscriptions and immutable mobiles remain attached to their original contexts and how they enabled traceability - that is, the capacity of reconstituting their own history and biography. Writing about his expedition to the Amazon forest as an ethnographer observing science in practice, Latour asks:

> Obliged at all cost to maintain the traceability of the data we produce with minimal deformation (while transforming them totally by ridding them of their local context), we would have seemed exotic to the indigenous people. Why take such care in sampling specimens whose features are visible only at such distance that the context from which they were taken will have disappeared? Why not remain in the forest? Why not go native? (1999, 47)

[15] For Latour knowledge derives from such movements and transformations, as well as that which remains and is maintained as constant throughout such transformations. 'Acts of reference', as he calls them, rely not so much on resemblance or representation as on a 'regulated series of transformations, transmutations and translations' (1999, 58).

For Latour, holding inscriptions or data constant, combined with the necessity for their simultaneous mobility, requires a trade-off between what is gained from transforming the forest into data – defined as a process of amplification – and that what is lost – defined as reduction. Reduction entails the loss of 'locality, particularity, materiality, multiplicity and continuity', whereas amplification presupposes the gain of 'compatibility, standardization, text, calculation, circulation and relative universality' (1999, 71). This entails that inscriptions and immutable mobiles, as they circulate, preserve some relations to the settings from which they emerged while other relations become necessarily lost. Latour however also notes that it is not only important to look at the cascade of inscriptions[16] which emerge as part of scientific work in itself, but also at the staging of the settings in which inscriptions are displayed and visualised. His work also pays attention to such stage-ings of settings, referred to as the 'staging of the scenography' in a rather Goffmanian fashion, that enables attention to be focused on 'one set of dramatized inscriptions' (1986, 17). It could be argued therefore that this focus on staging implies that inscriptions and immutable mobiles are disciplined and made visible in particularly orchestrated ways by their settings. In his ethnographic work in a scientific laboratory, Latour (1999) describes processes of entextualisation of inscriptions - relationally distilling them from their original contexts - as well as a process of

[16] Latour went on to internalise the criticism made about the overestimation of inscriptions in his arguments, claiming: 'Knorr has criticized this argument by taking an ethnomethodological standpoint (1981). She argues, and rightly so, that an image, a diagram, cannot convince anyone, both because there are always many interpretations possible, and, above all, because the diagram does not force the dissenter to look at it. She sees the interest in inscription devices as an exaggeration of the power of semiotics (and a French one at that!)' (1986, 21)

recontextualisation - incorporating them into stage-ings that orchestrate their aboutness.[17]

Conclusion

In this Chapter, I have briefly traced the conceptual development of the notion of *frame* in the bio-cybernetics proposed by Bateson (1955, 1972); the cognitive theory on frames developed for the field of artificial intelligence by Minsky (1975); the adoption of the term *frame* in the field of sociology by Goffman (1974); and its later transformation as part of STS and ANT in the work of Callon (1998). I have shown how the development of the concept of *frame* within these disciplines is tightly linked to the problematisation of contexts as given - to objects, interactions and phenomena in the world. What all four accounts show is a concern with how contexts are emergent, rather than *a priori* separated and bounded from interactions and cognition. Contexts are accomplished; their assemblage is important as this contributes to defining what is relevant for a given interaction or form of cognition, whilst shaping the aboutness of situations, objects and events. Bateson (1972) looked at how communication, in particular the interplay between signs and signals, helped to define the context of human and animal interaction. It could be argued that for Minsky (1975) on the other hand, the accomplishment of context was conceived of as effected and redefined whenever a form of human cognition - understood as a predefined

[17] Of course aboutness in the Amazonian and laboratory settings has to do wit hthe objectivity of phenomena, whereas in Goffman's account it is a concept designed to account for the organization of experience.

template frame - interacted with and was tested against a range of different situations. Finally, Goffman (1974) further expanded the notion of frames by conceiving of them as containments that can always become recontained through rekeyings and fabrications. As I have shown, in Goffman's account, a frame can serve as the pattern that forms the basis or substrata of further reframings or laminations which define the aboutness of social situations. Whilst these conceptual developments point to the establishment of contexts or frames via communication and social and cognitive interactions, the Chapter has also drawn attention to how the introduction of concepts like *device* and *overflow* in the work of Callon (1998) has been important to further expand the conceptualisation of frames.

To conclude this Chapter, I would like to highlight the relationship between overflows and data. Following Callon (1998), it could be suggested that the framing or the staging of data is always precariously sustained and never permanently closed. This implies that the latent relations to the setting form which data was extracted can always become reactivated or new ones established. In other words, the entextualisation and recontextualisation of data as an entity which is separate from the original interactional frame from which it was extracted requires a constant investment in framings and framing devices, a process which is nevertheless always at the risk of being capsized by overflows. Some recent developments in scholarship that examine the role of databases and data in everyday life have suggested that data is never fully separate from the contexts and objects to which it related in the first place (Lury et al. 2012; Fuller and Goffey 2012a). It is argued that as a consequence of the 'operation of dynamic feedback loops and the extension of their significance by their use in diverse, iterative and automatic information processing systems' (Lury

et al. 2012), the framing of data or inscriptions is becoming less static and bounded. In other words, the boundary between databases and data, and what becomes staged as their outside, is now even more precariously sustained, particularly if we also think of data as performative or reactive (Espeland and Saunder 2007).

Fuller and Goffey (2012a) claim, for example, that one of the crucial characteristics of databases is that they enable the establishment of productive and generative associations between the rules and structures they command, and the 'outside world'. These scholars suggest that the capacity of databases for ordering 'enables the introduction of combinatory and sorting for things outside the database as well as within it (…)'; but it also enables 'the production of new rules for relations, the harvesting of relations generated by the population of databases by live data generating processes' (2012a, 326). What is distinctive in Fuller and Goffey's account is that the nature of the relations that databases establish with their outside world are conceived not in terms of a logic of representation, where databases and the data they are comprised by establish unidirectional relations with particular entities in the world. They suggest instead that a more interesting way to think about databases is to understand them as topological machines, as devices that intentionally or unintentionally engineer connections between things, generating continuities but also disruptions (2012a, 326). Fuller and Goffey maintain that

> Any table of related data, a monograph or even a bus timetable, establishing links between a finite number of stable, discrete, and interconnected entities, is a topological machine. They establish networks of relations, the points of intersection between data and what data links to and triggers.' (2012a, 326)

Fuller and Goffey (2012a) also address the politics surrounding database framings and the constraining orders these afford. They see the normalised relations established by tables as particular kinds of orderings, orderings which are limited and hence political - but which are not immutable. Instead, they envisage that changes in the way databases order and structure the world are an ever-present possibility, perhaps entailing the disruptive establishment of 'false continuities, abrupt changes of person and personality, status, credit rating, security risk and so on' (2012a, 328). What is more, they see these events as starting points for the empirical investigation of databases - or what they also define as 'the arbitrary topologies of the laboratory of the world' (2012a, 328). In the following methodological Chapter of the thesis, I further explore the provocative suggestion that overflows and moments of data and database disruption present empirical openings and opportunities to investigate the framing of data and the work that different devices might do in this respect.

Chapter Two

Data Leaks as Empirical Occasions

> The leak then is an attempt to capture and mobilize the dynamics of unintended consequences, to enter into the domain of the accident, the double agent, confusion, and to render it fruitful. The leak, however, is never simply bivalent. For some, everything cries out for it, trying, if only by dint of time, to edge its way past the demon of the chamber of equivalence in which it is trapped, hungry for a connection, the accident of knowledge, for that or the chance to turn to dust, unnoticed and indifferent. (Fuller and Goffey 2012b, 103).

Introduction

In the previous Chapter I have delineated the conceptual lineage of the notion of *frame* as it forms part of different disciplines. In the following Chapter, I want to concentrate on the methodological implications posed by moments or occasions of *data overflow*. Overflows can be of different types and kinds. Callon (1998) in particular identified a wide range of events as overflows, including natural disasters, accidents, and - of particular interest to him - the production of positive and negative economic externalities. This Chapter argues that data leaks can also be understood as a type of overflow as conceptualised by Callon (1998). But rather than explaining leaks simply as overflows, the intention of my research project has been to take

advantage of leaks as 'empirical occasions' (Marres 2013); that is, as moments in which the framing of data becomes visible and the workings of framing devices more legible for social analysis.

Let me start here by describing how the research presented in this thesis has been designed around the configuration of two case studies and the different elements each comprise. The research began with the identification and description of two different data leaks, one from a commercial and one from a public national database; this was followed by the collection of materials (see below) and the analysis of the devices which surfaced as a result of these different data overflows. The objective was to design and delimit two case studies based on these materials and the framing devices that the leaks drew attention to. The first case study, entitled '*The Making and Unmaking of Personal Data*', looks at the so-called 'AOL data-leak' that took place in August of 2006 in America. This case study engages particularly with the unintended public release of databases containing nearly 20 million anonymised search queries typed by America Online (AOL) users over the period of three months. The case study presents a description of the leak and a depiction of the personalising functions of search engines. It also compiles a range of materials that have emerged and have become visible as a part or a consequence of this particular data overflow - including an email; an art documentary; the leaked database and data units themselves; a journalistic exercise undertaken by The New York Times; and data research regulatory frameworks. This case study focuses on a range of devices and techniques that are deployed for framing data as personal or impersonal, including reidentification demonstrations; sequential data arrangements; informed consent and anonymisation.

The second data leak that the thesis examines is one that was part of a controversy around the Office of National Statistics and Census in Argentina at the start of 2006, and it comprises the second case study, entitled '*Price Formation and Statistical Compositions*'. This particular data leak revolved around different requests made by government officials to release granular and disaggregated data pertaining to the measurement of the Consumer Price Index (CPI) in Argentina; and the deployment of the Statistical Secret Act that prevented such disclosure. In this case study, I start with a brief description of the data leak and the problems it raised in relation to the stabilisation of price as 'pure' in the measurement of CPI. I assemble and analyse a range of materials that have emerged as part of this particular data overflow – including legal warrant documents [Legajo de Causa Fiscalía de Investigaciones Administrativas 2007; Pedido de Indagatoria por parte de la Fiscalía de Investigaciones Administrativas 2007]; seven interviews I undertook with actors involved in the dispute over the disaggregation of data; two investigative journalism books dedicated to the controversy; and three alternative inflation indicators that emerged in response: CPI Congress, The Billion Price Project and Premise. The case study draws attention to techniques and devices that participate in the formation and stabilisation of prices for the measurement of inflation, including identification procedures; aggregation; digital scraping and digital imaging.

Before turning to each case study, in this Chapter I describe the methodological assumptions that informed the configuration of these case studies. The Chapter is divided into three main sections. The first section describes what it entails to consider data leaks as an inventive method amenable to repurposing for social scientific

objectives. The Chapter then moves on to examine, also from a methodological standpoint, what is involved in defining leaks as a type of societal experiment. The last section justifies the research design which treats data leaks as means to study data frames and the devices that enable these, rather than defining leaks as an object of study in themselves. The Chapter suggests that one way of empirically taking advantage of society's experiments such as leaks is to configure them as cases, an operation described by Charles Ragin (1992) as *casing*.

Leaks as Inventive Method

I want to start by briefly drawing on the notion of inventive methods or devices (Lury and Wakeford 2012). A range of recent accounts[18] in the social sciences have explored and problematised the potential of the notion of *device*[19] to elicit alternative and critical perspectives on the effects provoked by the use of different methods and measures. The development and deployment of the term *device* from a methodological perspective can be seen as part of a series of recent conceptualisations in the social sciences that have addressed the politics and

[18] See for example Ruppert et al. (2013).

[19] Apart from drawing on Callon et al.'s (2007) notion of device, Lury and Wakeford (2012) trace the definition of the concept to Michel Foucault as in *dispositif* or *apparatus* (1980) and Giorgio Agamben who claimed that 'further expanding the already large class of Foucauldian appratuses, I shall call an apparatus literarly anything which has in some way the capacity to capture, orient, determine, intercept, model ,control, or secure the gestures, behaviours, opinions or discourses of human beings' (Agamben, cited in Lury and Wakeford 2012, 8).

ontological capacities of methods deployed for social research.[20] The introduction of the term *device* to think about method has been particularly important for Celia Lury and Nina Wakeford (2012) because it takes account of a number of things. Firstly, it recognises that methods, tools or practices of knowing the social world are not necessarily used always and everywhere in the same way, but function differently and are productive of different effects depending on the configurations that they become part of and are attached to (2012, 9). On the other hand, device is a term that has been useful in recognising that 'knowledge practices, technical artefacts and epistemic things are encoded in everyday and specialized technologies and assemblages in which agency is not the sole privilege of human actors' (2012, 9). Also importantly they note, 'the notion of device is welcome insofar as it draws attention to the semiotic-material relational-doing-thingness of methods' (2012, 9).

Lury and Wakeford's edited book on inventive methods can be read as a collection of devices or tools – such as anecdote, category, configuration or experiment - which are explicitly oriented towards 'investigating the open-endedness of the social world' (2012, 2). For Lury and Wakeford, methods do not describe an already static and predefined social world from the outside, but enable 'the happening of the social world – its ongoingness, relationality, contingency and sensuousness – to be investigated' (2012, 2). Importantly, they also insist that it is not possible to apply a method or device as if it were 'indifferent to the problem it seeks to address, but that method must rather be made specific and relevant to the problem' (2012, 3).

[20] See for example Law and Urry (2004); Law (2004); Adkins and Lury (2009); Adkins and Lury (2012); Back and Puwal (2013); Ruppert et al. (2013); Law and Ruppert (2013); Savage and Burrows (2007).

Inventive methods are distributed. This implies that scientific and academic research do not have a monopoly on methods, and that methods are deployed in a number of different settings by professionals and laypersons for different purposes. Inventive methods are hence implicated in the '(re-) organization of processes of knowledge making' (Marres 2012, 3).[21]

I want to suggest that leaks or overflows could also be considered as a form of inventive method or device. Matthew Fuller and Andrew Goffey (2012b) have already implicitly described leaks in this way by claiming that they might constitute a form of epistemic intervention capable of organising knowledge (a)symmetries. They state:

> [A leak] operates by means of establishing relations between different registers of the inside and outside, between those who already know, those who might possible know if they had the time, those who should have been told, the

[21] Lury and Wakeford (2012) go on to describe three other important aspects of inventive methods or devices. First, they claim, inventive methods are detachable from specific problems or situations and become usable in a range of contexts, being continually introduced into new ones (2012, 10). Furthermore, inventive methods always operate as part of a configuration or apparatus 'modified in specific uses, undergoing transformation (…) in relation to particular situations, particular problems' (2011, 11). Secondly, inventive methods also open up new relations between the specific and the general. Such devices have for Lury and Wakeford (2012) a multiple capacity for generalisation – to show up, circulate and proliferate - in a way that that is not necessarily universalistic. Finally, inventive methods also come to 'recognise the properties of the medium – or media – of methods for whatever inventiveness happens' (2012, 18). This presupposes that the material and semiotic entities that bring methods alive should also be taken as agents in the research process, and that the material agency of methods 'should not be seen as a delegation but as a translation' (2012, 18).

> scandalized, doomed or grateful recipient of truth, and those who do not need to know or would rather not, among others. Thus the particular style of the leak resonates with all of these figures, each of whom may adopt one or more of the available positions according to the nature of the leak and their possible operations on and through it. (2012, 103)

Fuller and Goffey also claim that although leaks have become a generalised way of disseminating and staging truth claims, forms of knowledge and a range of other narratives, a general taxonomy of the leak understood as media is yet due to be developed (2012b, 100). In an attempt at identifying the ways in which leaks arise and make themselves amenable to analysis, Fuller and Goffey (2012b) describe a broad range of different events and occasions as forms of leaks.[22] Leaks, they claim, should be understood as the staging of a particular genre through which narratives become formatted, publicised, and circulated - but also forgotten. They suggest that leaks be understood as a form of media that might or might not be 'received by an apparatus capable of distinguishing it from another noise' (2012b, 103). However, leaks arise, make themselves amenable and become mobilised, visible and formatted for different purposes in a variety of different ways. Leaks can be staged as forms of decoy, as ways of 'revealing and handling materials before they can actually said to have occurred' or again as mechanisms for establishing 'relations between different registers of the inside and outside' amongst an array of other possibilities (2012b, 101). In other words, whereas leaks are never considered as belonging to or forming

[22]For example the controlled release and revealing of draft documents; the manipulation of off-the-record information; an accidental slip of the tongue; an uncalculated and unwitting cc; a lost briefcase, etc. (2012b, 100-103).

part of an - already framed - phenomenon, Fuller and Goffey (2012b) draw attention to how they are nevertheless bounded and staged as specific events or occasions in themselves.

Michel Callon (1988) has also addressed the specificity of the onto-epistemic interventions enabled by different types of overflows, but with a particular focus on scientific research and measurement. Callon (1998) claims that overflows emerge from the efforts undertaken by scientists and experts to measure them and to render their consequences calculable. Callon (1998) understands such measuring procedures as facilitating the reframing of interactions that gave rise to the overflow in the first place. In other words, the techniques involved in measuring the overflow are seen as interventions that open up a space of possibilities for externalities to become internalised. It is worth noting that this is a point that has particular methodological and inventive importance: measuring instruments and their capacities open up spaces that enable the different agents involved in the definition of an overflow or leak to either negotiate an agreement, calculate their respective interests or do both (Callon 1998, 259). Callon further argues that 'the possibility and viability of this negotiation both depend on the availability of instruments capable of producing incontrovertible measurements' (1998, 259) and he suggests that this is a point where both natural and social sciences can make a contribution - by reframing interactions and, by means of measurement, allowing externalities to become internal to, or remain outside, any given transaction. Importantly, Callon also claims that 'these investments apply and produce both knowledge, in that they cause hitherto invisible links to appear, and also a reconfigured collective in which these now visible and calculable links have been renegotiated' (1998, 259-260).

It can therefore be argued - following Callon (1998) and Fuller and Goffey (2012b) - that leaks have different functions and capacities. Firstly, they are enacted by scientists through annotating and measuring procedures which enable the inventive renegotiation of what is included and left out as part of different frames. Secondly, leaks can also be deployed in a range of different settings – including academic research - as a way of handling and distributing material that has not yet occurred (Fuller and Goffey 2012b); or as a device of (a)symmetrical knowledge distribution, as for example when a draft is circulated in advance of the actual publication of a scientific article. But as Fuller and Goffey (2012b) also suggest, leaks or overflows are methods generally deployed and staged as part of public, everyday and digital life without necessarily having to rely on specifically scientific measuring procedures to enact them.[23] Considered as such, leaks constitute a distributed method, always enmeshed in different configurations, being deployed for diverse purposes and actors, and having numerous and sometimes unexpected effects.

Because some leaks are formatted, relatively successfully, to be detectable by different apparatuses as visible forms of noise, they also prove to be generative of a diverse range of materials and effects amenable to social scientific analysis. This provokes however the following question: how to deploy and handle them for the

[23] For a visualisation and account of data leaks, breaches and hackings see http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/, last accessed 7th August 2015.

purposes of social scientific research?[24] My thesis takes data leaks as 'empirical occasions' (Marres 2013), while also proposing a reorganisation and repurposing of the diverse range of materials and effects which have emerged as a consequence of the staging of data leaks as part of public and digital life. By taking data leaks as empirical occasions and by reframing and redeploying them for the purposes of social research, my aim has been to develop a research design that is relevant and specific to the social study of data frames. In particular, I have been concerned to develop a research design that is capable of making visible the effects and capacities of framing devices for the articulation of data's ontology and political economy in contemporary, everyday settings. In short, I have assembled the different materials and effects enabled by the leaks, and repurposed them to configure a suitable context in which questions about different data framings can be conceptually addressed and worked upon.[25]

One of the distinctive contributions that the following leak-enabled social research study proposes is to accomplish the repurposing and reordering of materials and effects by considering two other distinctive inventive methods: that of the *experiment* and that of the *case* (study). In what remains of the Chapter I unpack some of the assumptions that underpin the selection of the research design, by analytically

[24] A brief note on the 'repurposing of methods' is due: as Marres and Weltevrede have argued, repurposing methods, techniques and materials that are exogenous to social research tends to involve 'the importation of categories and formats into social research that are, strictly speaking, external to it' (2013, 315). Rather than seeing this as a limitation, Marres and Weltevrede (2013) have argued that although exogenous formats present particular complications for social research they also have distinctive affordances.

[25] Social scientists have relied on similar occasions with a similar purpose, including the analysis of controversies, accidents, disasters and transgressions.

engaging with a range of social scientific traditions that have also inventively problematised and deployed *experiments* and *cases*.

Experiments as Inventive Method

There are, as Fuller and Goffey (2012b) have noted, many types and kinds of data leaks, and if we are to understand them as an inventive method, it would be reasonable to argue that the function they fulfill and the types of effects they achieve vary depending on the different configurations and apparatuses by and through which they are staged. In what follows, I would like to consider the two data leaks that the thesis explores as types of *society's experiments*, as conceptualised by Mary Morgan (2013). Society's experiments are different to laboratory experiments. The latter are usually characterised as 'controlled interventions into nature designed and executed by scientists' and carried out in artificial environments that shield 'the objects and events of the experimental materials from the effects of other factors and disturbances by strict protocols of both intervention and control' (Morgan 2013, 324-343). In other words, laboratories are protected spaces which, because they are stripped of 'contextual complexity and environmental variation', are 'breeding zones for strange, unforeseen and unapproved knowledge, skills and techniques' (Schwarz and Krohn 2011, 126). Laboratories instigate a boundary between a controlled inside and an

uncontrolled outside, creating an inside that enables actions to be rendered inconsequential (Guggenheim 2012, 2).[26]

Laboratory experiments are usually, in STS literature, contrasted to 'field experiments'. The latter are experiments 'designed and carried out by scientists to ape the laboratory conditions in the field' (Morgan 2013, 343). Scientists intervene in the real world, controlling these interventions by deploying multiple experimental design features such as randomised control groups – the clinical trial being perhaps paradigmatic in this respect. But as Morgan also notes, 'other kinds of experiments also occur outside of the laboratory, in a variety of research sites, in different forms, and in which different elements of the laboratory's ideal type hold' (2013, 343). Morgan conceptualises one of these types of experiment occurring in the field as, precisely, society's experiments. These types of experiments are for Morgan 'natural situations holding the characteristics of experiments' (2013, 343), and in contrast to traditional field experiments they occur outside the control of (social) scientists, and without any deliberate intervention from them. These events, Morgan argues, occur 'by experimental intervention (whether by random accident, natural/social causes, or known human intervention) within an open environment' (2013, 344). One of the distinctive characteristics of society's experiments is that their controls and interventions become instantiated 'naturally', that is, without the involvement of social scientific work.

[26] Social scientific literature on laboratories and experimental methods in different disciplines abounds, and this Chapter is not intended to cover the wide range of conceptual perspectives on this subject. For a recent and compelling account of the role played by experiments in the social sciences see Guggenheim (2012).

Morgan identifies four different sub-types of Society's experiments,[27] classified according to the ways in which the 'ceteris paribus controls of the laboratory can be understood to be present' (2013, 344). Let me here concentrate on the third sub-type, as this is the one most relevant for my argument. Morgan identifies the third sub-type of societal experiments as 'events that stand out in some way as unusual and yet take place in very stable situations' (2013, 345). Morgan further argues:

> Many, perhaps most, events in the natural and social world occur with lots of other events happening around them, but some individual events happen in short time periods in specific places where it is reasonable to suppose that the environmental features are very stable, and the other causal factors (that might normally vary over space or time) are also rather stable. For example, Orson

[27] Here, I am only describing the third sub-type described by Morgan. The first type is defined as 'massive interventions that makes issues of control and isolation from disturbing causes irrelevant', and, she continues, followed that 'there are events or processes of sufficiently massive scale and spontaneity creating very substantial effects that swamp any and all other effects or disturbances from other causes and from instability from the environment' (2013, 345). Examples of this type would be epidemiological, large- scale environmental catastrophes and social revolutions. Morgan defined the second sub-type of society's experiment as 'situations of total isolation, where natural or social processes go on in an uncontrolled way with their local (isolated) environment' (2013, 346). Examples of these would be where isolated societies are formed due to constraining circumstances such as war, or societies which go on to develop particular characteristics due to such isolation or constraint. Finally, the fourth sub-type is described by Morgan as 'situations where the environment is controlled and within which Society's experimental interventions are also carefully controlled' (2013, 346); she explains that 'some situations present themselves as if an experimenter had designed a laboratory experiment within the world' (2013, 347). Examples of these she suggests are cases of 'child prisoners' or 'wild children' on one hand, or those of 'total institutions' as identified by Erving Goffman on the other (Morgan 2013, 347).

> Welles's radio broadcast of H. G. Wells' War of the Worlds that caused such panic in America in 1938 happened in such a short period that there was no reason to assume the environment and other factors that influenced people's response to this were not stable. This self-contained event was effectively treated as one of Society's experiments: it prompted a sociological (interview-based) investigation into how people responded and why certain people panicked (…). (2013, 346)

Morgan also argues that this type of experiment, involving 'the presence of an intervention with unambiguous effect with a high level of stability in the environment and in other influencing factors' (2013, 347) is the most prevalent form of Society's experiment. But more importantly, she further argues, these events happen in such a way that they can be 'studied by scientists after the event as if they were designed experiments in the field' and where 'no retrofitting of the events nor of the environment is required' (2013, 347). Morgan uses another sociological example to illustrate the uses made by social scientists of this sub-type: namely, that of Robert Merton's investigation of a mass propaganda event.[28]

I will elaborate on how my own research has taken empirical advantage of such events later, but here it is worth emphasising that society's experiments generally involve events or occasions unfolding as part of - or staged by – actors in society at large, rather than caused by the controlled interventions of scientists guided by rules of scientific method. If read in those terms, society's experiments resonate with the more radical ideas developed by the early Chicago school of sociology with regards

[28] See Merton et al. (1946)

to 'societal experimentation'.[29] Albion Small, one of the main exponents of this early tradition, suggested that

> All the laboratories in the world could not carry on enough experiments to measure a thimbleful compared with the world of experimentation open to the observation of social science. The radical difference is that the laboratory scientists can arrange their own experiments while we social scientists for the most part have the experiments arranged for us. (Albion, cited in Gross and Krohn 2005, 67)

Basing the idea of experimentation not in scientific method but in social life itself, 'and therefore importing the experimental design from the object under study into the method of the sociological observer', is indeed a remarkable move (Gross and Krohn 2005, 67). The notion of a self-experimental society or the idea that society itself is operative in designing experiments (Gross and Krohn 2005) has also been recently elaborated by Noortje Marres in her discussion of 'sustainable living experiments' (2012). Marres argues that these types of experiments are based on the 'modification of habits and habitats according to a fixed procedure: they are a way of implementing changes in everyday routines and living spaces according to a protocol' (2012, 78). She suggests that sustainable living experiments constitute a proliferating media form in that their circulation participates in 'the replication of sustainable living, or versions thereof, across social life' (2012, 78).

[29] See Gross and Kohn (2005).

Marres claims that experiments conceived of as devices or inventive methods[30] serve a broad range of objectives all at once, and that they therefore are 'multifarious instruments' (2012, 82). Public experiments for example perform at least two functions simultaneously: on one hand, they facilitate the introduction and domestication[31] of techno-scientific objects as part of the fabric of everyday life; on the other hand, as sites that enable 'a politics of contestation', public experiments can also be thought of as undertaking social or political work, and as mechanisms for enrolling or enlisting actors in technical or scientific projects (Marres 2012, 80). Experiments are also sometimes deployed as a 'genre of publicity' (Marres 2009, 2012) and as Harry Collins (1998) has also noted in this respect, there are substantial

[30] Marres's (2009) account of sustainable living experiments is included as part of Lury and Wakeford's collection of inventive methods, in which experiments feature prominently. While my account here draws on the work of Marres (2012) mainly, it is worth briefly noting that in the same collection, Steven D. Brown (2012) too focuses on understanding the psychological experiment as 'one site of activation of a device for dramatizing human conduct that is distributed throughout a variety of cultural and historical settings' (2012, 72). He notes that the experimental device in the cases he analysed involved on the one hand the orchestration of an 'abstract-realist version of a matter of concern that can be nevertheless treated as a functional analogue'; whereas on the other hand, the experimental device also pursued a 'pure expression of some essence by multiplying, repeating, juxtaposing and reducing everyday elements until they take an entirely new form' (2012, 72).

[31] The concept of domestication has been used in ANT to account for the movement of objects into and within existing socio-technical arrangements and how these movements change the configuration of such socio-technical arrangements too (see Berker et al. 2005).

differences in the effects produced when the trying, showing or displaying of an experiment takes hold.[32]

Marres attributes a further capacity to experiments also here worth noting here – namely that of rendering phenomena visible and accountable. By taking an art event as an experiment, for example, she argues that the introduction of an art device into a public space could be read as a mechanism to render a phenomenon 'describable and thereby accountable' in particular ways (2012, 88). For Marres, experiments enable the 'rendering accountable of everyday objects, settings and action – something that ethnomethodologists and others have identified as an important accomplishment of social research' (2012, 90). Indeed, in both interactionism and ethnomethodology, experiments have been used as mechanisms for deducing how expectations are organised by way of frames or social interactions otherwise backgrounded and taken for granted.

I would like to refer here briefly to interactionism and ethnomethodology again, but this time in order to focus on how these traditions have used experiments as devices for displaying that which was deemed to be 'social' in very specific ways. Interactionism and ethnomethodology are two sociological traditions that have

[32] Collins (1998) notes that 'in mid to late seventeenth century England there was a linguistic distinction … between 'trying' an experiment and 'showing' it … The trying of an experiment corresponds to research proper, getting the thing to work, possibly attended with uncertainty about what constitutes a working experiment. Showing is the display to others of a working experiment, what is commonly called a demonstration…. I want to say that trying was an activity which in practice occurred in relatively private spaces, while showing [was an event] in relatively public space. The career of experimental knowledge is the circulation between private and public spaces' (Shapin, cited in Collins 1988, 727).

performed experiments as either natural or controlled forms of disruption in the normal flow and continuity of frames or interactions. These disruptions were taken as occasions in which a property or element within ordinary activity was displayed and contrasted in a clarified and clarifying way, either by 'flooding in' or 'flooding out' (Goffman 1974). As discussed in the previous Chapter, Goffman (1974) described these occasions as moments in which frames break by either becoming flooded in or flooded out by a range of elements, behaviours or improper involvement which do not necessarily belong to or are not accepted by other agents as being part of a given frame. Goffman claimed that

> Given that the frame applied to an activity is expected to enable us to come to terms with all events in that activity (informing and regulating many of them) it is understandable that the unmanageable might occur, an occurrence which cannot be effectively ignored and to which the frame cannot be applied, with resulting bewilderment and chagrin on the part of the participants. In brief, a break can occur in the applicability of the frame, a break in its governance. (1974, 347)

Goffman (1974) described 'flooding outs' as occasions on which it is not possible to reinsert elements of a framed event into appropriate involvement or bring it back under the control of the frame. He treated these as situations which provoke the limits of the frame to become visible; as such, he understood them as 'natural experiments' in which that which is implicit and concealed about the organisation of the frame and what it enacts, enables and constrains can be unpacked, unraveled or revealed (1974, 564). Goffman used the intervention of an intermediary and foreign element as a way

of empirically explicating the organisation of the frame itself, and also as a means of describing what the frame was able to contain and what was excluded from it. Goffman therefore took advantage of what he also understood as empirical occasions in which frames are made visible via the elements that inflow or outflow as part of a given interactional dynamic.

Whilst part of Goffman's experimental method was deployed to observe, identify and describe what he called natural experiments involving 'flooding in and outs', Harold Garfinkel (1967) on the other hand proposed that social scientists could also engage in provoking overflows via 'controlled experiments'.[33] The breaching experiments proposed by Garfinkel and his disciples entailed the controlled disruption of ordinary scenes[34] (Marres 2012). Garfinkel claimed with regards to such experiments that they

> […] attempt to detect some expectancies that lend commonplace scenes their familiar, life-as-usual character, and to relate these to the stable social structures of everyday activities. Procedurally it is my preference to start with familiar scenes and ask what can be done to make trouble. The operations that one

[33] It is also interesting that Garfinkel further noted that although the activities he and his disciples undertook carried procedural emphasis, they should not properly speaking be conceived of as experimental. He noted that these disruptive exercises were more like demonstrations designed as 'aids to a sluggish imagination' (1967, 38).

[34] For example, students were instructed to engage in ordinary conversation with acquaintances in an unusual way by continuously insisting that persons clarified the meaning and sense of their commonplace remarks (1967, 42). Interestingly too, in one of his scattered methodological notes relating to breach experiments Garfinkel noted that the knowledge derived from disrupting the structures of everyday activities and how they are routinely produced should fold back into how sociologists proceed in the subsequent design of 'the effective production of desired disturbances' (1967, 38).

> would have to perform in order to multiply the senseless features of perceived environments; to produce and sustain bewilderment, consternation and confusion; to produce the socially structured affects of anxiety, shame, guilt, and indignation; and to produce disorganized interaction should tell us something about how the structures of everyday activities are ordinarily and routinely produced and maintained. (1967, 37)

By disrupting the normal flow of conversation usually undertaken between two persons as part of their everyday activities, for example, Garfinkel proposed to elicit the background of unnoticed features of common discourse which is used to 'entitle persons to conduct their common conversational affairs without interference', and he suggested that 'departures from such usages call forth immediate attempts to restore a right state of affairs' (1967, 41-42). Via such experiments Garfinkel also sought to demonstrate 'the kinds of expectancies that make up a '"seen but unnoticed" background of common understandings, and how they are related to a person's recognition of stable courses of interpersonal transactions' (1967, 44).

In the two types of experiments described above, interventions were used as a means for making visible and explicating the structure, organisation and dynamics of the frame being experimented upon. What the experimental intervention enabled - in particular via the observation or introduction of disruptive elements - was the frame to become front-staged and thus rendered available for analysis. In these traditions the objective of the experimental device is to render frames empirical; in other words, the experimental device was also used as a technique for the [partial] production and display of social reality (Lezaun et al. 2012, 279). The experiments 'activated the

latent energies of everyday life', and in so doing revealed and displayed a new reality (Lezaun et al. 2012, 280). However, for the experiment to succeed in revealing these realities, interventions needed to be handled and staged in precise ways. Lezaun et al. referred to this double force of the *laboratory experiment* as a form of 'provocative containment' (2012). Lezaun et al. (2012) suggest that we should not think of these techniques as 'imperfect methods to re-present social phenomena accessible elsewhere or by other means' (2012, 280). Their approach instead is to account for these techniques as mechanisms for producing 'vivid and otherwise unavailable renderings of social reality', and to understand them as capable of enacting social reality 'in a particular demonstrative and explicit form.' (2012, 280).

Data Leaks as Society's Experiments?

So in what sense (other than metaphorically)[35] can the data leaks that are the focus of this thesis' empirical investigations be identified as a type of society's experiment? And if we do approach these leaks as society's experiments, then what kind of function do they play and by which means do they also become useful for social scientific purposes? My thesis empirically engages with the leak of data from two rather different settings, and traces how this leaked data was subsequently reframed. The first case study that the thesis delineates is based on the so-called 'AOL data leak'[36]. This leak consisted of the public release of search keywords in a database format, which was later used by artists, journalists and activists to attempt to identify

[35] See Guggenheim (2012).

[36] See https://en.wikipedia.org/wiki/AOL_search_data_leak, last accessed 21st August 2015.

and profile the authors of such searches in their everyday settings. The second case study is drawn from a similar event: the disaggregation of a national statistics database to the point where data that was normally legally protected in fact enabled the identification of the everyday products and settings that produce the measurement of a national inflation indicator. In both cases, one of the most important effects unleashed as a consequence of the overflow of data was reidentification, by means of data of persons, objects and contexts.

It could be argued that these two specific events resonate with Morgan's (2013) third sub-type of society's experiment described above. To begin with, they could be characterised as unusual or atypical in two respects. Firstly, search keywords are configured as expressions of information needs and intentionality by search engines. In the AOL leak, however, search keywords were mobilised to identify and qualify persons in other particular ways instead. In the case of the national inflation indicator, on the other hand, prices and product characteristics are organised and formatted to index the value of consumer goods in relation to practices of everyday shopping. The inflation leak however attempted to identify product characteristics and prices to reveal different socio-demographical characteristics of the settings and products themselves. Secondly, both of these leaks were unusual events because they made data visible to consumers and online users in ways that are not normally possible. While users can see what search keywords they type, they are unable to configure the history and serialisation of their searches in any other meaningful way. Similarly, whilst there are many available techniques and devices which enable consumers in everyday shopping settings to assess price fluctuations, the particular products and prices taken into account to measure inflation - those that ultimately provide an

objective reference of price variation - are usually legally prevented from being accessible and visible to consumers and citizens as part of their everyday shopping practices.

Most importantly, in terms of the classification of society's experiments provided by Morgan (2013), it could also be argued that in both cases the data leaks triggered an intervention with unambiguous effect: that of re-purposing data to reveal both specificities and contexts - ranging from personal and subjective traits in the case of search keywords to socio-demographic characteristics of both consumer goods and settings in the case of inflation data. Furthermore, the leaks were not necessarily guided by scientific methods or by the work of scientists but by a different range of actors in society. On top of this, both leaks and the identification processes they enabled could be equally considered as society's experiments if read as practices oriented towards the domestication of different forms of data – search keywords and inflation statistics – as part of everyday life. As noted above, understanding public experiments as socio-technical domestication devices has been a prominent argument in the STS tradition (Latour 1993; Collins 1998; Marres 2009, 2012). Finally, and also following the methodological approaches of social interactionism and ethnomethodology described above, one could argue that by flooding in and out and disrupting everyday framings, data leaks rendered certain aspects of the relations and dynamics they interfered and interacted with visible and accountable.

However, what Morgan (2013) shows when engaging with some paradigmatic uses of society's experiments in sociology and economics in particular is that these moments or events also serve as empirical occasions, or 'strategic research sites', amenable to

repurposing for other social scientific uses and objectives. Most importantly, they can be hybridised with other epistemic genres beyond the experiment. More often than not, however, and because there is no evident environmental stability to act as a form of control for most events that happen in the social world, scientists attempt to experimentalise such events further by 'reconstructing the ceteris paribus controls to turn a natural or social event into an experimental site for their study' (2013, 349). Morgan argues that 'social scientists reconstruct the normal events of life into natural experiments by post hoc "reverse designing" the natural/social situation in its environment into an experimental one' (2013, 349). Morgan describes three types of reverse design strategies usually deployed by social scientists: assigning or identifying randomisation; establishing control versus treatment site groups; and the use of statistical analysis as a form of control.

Another way in which social scientists tend to experimentalise society's experiments is by isolating the event from the interferences that other related events might impose post hoc. This type of mechanism or technique provokes what Morgan defines as 'social scientizised experiments' (2013, 349). Morgan (2013) however is particularly uncomfortable with the execution of isolation procedures. She argues that laboratory, field and retrofitted experimental designs proceed by focusing on one aspect or relation, and are deliberately designed to 'delete' or otherwise control all other relations. She argues that such designs tend to be reductionist as they stabilise and control one set of relations only and, in the case of retrofitted experiments in particular, this certainly does not do justice to the generative capacities of society's experiments, understood as empirical occasions. She argues that society's experiments should be taken as rich sites for social scientists to take advantage of

empirically because they allow access to 'concentrated moments and effects' (2013, 355) that, if redeployed inventively, can trigger further experimental effects - effects that do not necessarily require controls in an experimental laboratory sense. Morgan claims that some if not most events that unleash an experimental intervention will not be flexible enough for reconstruction or manipulable enough for post-experimental analysis in such a way as to isolate the effect of any one element in the web of relations in which it happens to occur (2013, 355). And in the cases where such reconstruction is feasible, it remains problematic and possibly counterproductive to reconstruct and reduce complex social events that unleash experimental effects into 'social scientizised experiments' (2013, 349) purely designed to isolate and control.

I therefore suggest that one way of empirically taking advantage of data leaks in a non-reductionist way is to reverse-design and *post-experimentally analyse them* by redefining them as *cases*. The empirical objects of study that my thesis delimits and attempts to stabilise are not however society's experiments or data leaks as such. Instead the thesis repurposes data leaks in the form of cases through which the framing of data can be analysed and conceptualised. The aim is to *open up and extend data leaks' affordances and generative effects* as a means towards exploring and rendering the devices and techniques that contribute to frame data in particular ways. I call this *casing*, after Ragin (1992).

In casing data leaks by analysing the materials that they generate and unleash, and the effects which become visible as a result, the aim of the thesis is to account for how the relations of data with a range of entities and phenomena become articulated. In short, I take data leaks as empirical occasions and openings that, when cased, provide

a range of materials and effects that facilitate the social study and conceptualisation of data frames. I have done this by mobilising a specific research design - that of the case study.

Casing Leaks

In their discussion and description of field and natural experiments, Astrid Schwarz and Wolfgang Krohn note that during biological experiments that deal with invasive plant species scientists intervene in the test site by bounding the plants or object of study 'with a red and white striped ribbon' (2011, 129). Whilst acknowledging that it is neither 'Nature' nor 'Society' which ultimately orchestrates the experiment, scientists can be influential in the demarcation of experiments as such and play a role in refracting, extending, concealing, revealing or renewing their salient features and effects. In the following section, I describe the particular binding/ribbon that I have deployed around data leaks – that of the case - and describe how the research activity of casing can also be put to use as an alternative form of 'provocative containment' (Lezaun et al. 2012).

It could be argued that as with many inventive methods, the case has been put to work, adapted and configured as part of a myriad of genres, deployed in a varied range of contexts, and mobilised by diverse apparatuses. Laurent Berlant stated in this respect that

> The case represents a problem-event that has animated some kind of judgment. Any enigma could do – a symptom, a crime, a causal variable, a situation, a stranger, or any irritating obstacle to clarity. What matters is the idiom of the judgment. This varies tremendously across disciplines, professions, and ordinary life scenes: law, medicine, universities, sports bars, chat shows, blogs, each domain with its vernacular and rule-based conventions for folding the singular into the general. Psychoanalysis mobilized the case-study genre to worry at questions of obscured causality, intention, and consent. Biopower uses the case study as a primary instrument in its machinery for making individuals into normative social units. It took aesthetic form in documentary and ficto-narrative genres (the detective story, the fictional autobiography, the medical mystery, the still life) and then in interpretative scholarship. It became available as an ordinary mode of life explanation, especially after the development of mass cultural norms of inducing identification. It took shape in the social sciences and business as a way of rationalizing and debating about how to manage singularity and generalization in research design. (2007, 664)

The multifarious capacities of cases are certainly very apparent within social scientific research: responses to what a case 'is' and what cases are cases of vary enormously across different methodological traditions (Ragin and Becker 1992). I will not go into detail here regarding the wide-ranging ontologies and epistemologies presumed in the multiple uses of cases in the social sciences,[37] but instead focus on describing the case approach that has been particularly pertinent to my research agenda.

[37] See for example Byrne and Ragin (2009).

My research design follows the conception of cases developed by Ragin (1992). Ragin suggests that cases - rather than being found, or naturalistically described 'out there' – are the outcome of a research tactic and activity he defines as 'casing' (1992). It could be argued that Ragin develops an empiricist conception of reality, broadly conceived of as 'limitless in its detail, complexity, specificity and uniqueness' (1992, 217). For him, conceptual and theoretical ideas provide ways of seeing the empirical world and structuring descriptions of this world. He argues therefore that in this light, 'empirical research could be seen as culminating in theoretically structured descriptions – understandings that result from the application of constraining ideas to infinite evidence' (1992, 218). For Ragin, theoretical descriptions are nothing more than 'remarkably feeble devices for structuring description – for generating the results of social science' (1992, 218). He further argues that the conceptual and the empirical are mutually dependent, and that social scientists transform empirical renderings into results with the aid of ideas whilst making sense of ideas and their relations by testing them with and against empirical materials or evidence. Casing hence is the product of basic research operations whereby the conceptual and the empirical inductively emerge as a dynamic and transient unit or whole. Cases are thus the outcome of specific limiting operations, and are a testimony to or an expression of the empirical world; but they are also an agent that contributes to thickening the complexity of that world. Casing configures a distinguishable mode of togetherness (Stenner 2012, 136), one that should be conceived of as a conceptually elaborated description of the empirical on the one hand, and simultaneously as an agent circulating and affecting the empirical on the other.

As an epistemic genre or a way of doing social research, cases have also been defined in terms of a series of other characteristics: firstly, they have been conceived of as a means to study a 'bounded whole object of analysis' composed of different parts or elements whilst retaining a 'considerable degree of open-endedness'; that is, cases facilitate the expansion or contraction of the boundaries that delimit a unit of analysis or object of study and its context or background throughout the research process (Morgan 2012, 668). In the social scientific tradition, cases have also demanded a 'considerable depth of engagement' with 'dense evidential materials' used in the analysis of a conceptually crafted whole. Case studies have also been described as enabling the deployment of different research methods, and their outcome is generally expected to be a 'complex, often narrated account' that typically contains, combines and organises raw evidence and analysis (Morgan 2012, 668).

Andrew Abbott (1992) further notes that in some branches of qualitative social research, cases tend to be associated with 'caseness', in the sense that they are indicative of the relevance of the processes and techniques used in defining them as such. Abbott (1992) argues that to establish 'caseness', research must account for appearance, disappearance, combination and transformation. Far from being problematic, these instances of the research process can be used as tactics to delimit, shape and display a case as a unit or whole. Indeed, one of the objectives of case-based social scientific qualitative research has been to arrive at a delimited case or set of comparable delimited cases as one of the main outcomes of the research process itself.

Morgan has also argued in relation to case studies that

> Social scientists, at various times and in various fields, have argued that case studies are not primarily vehicles for theory testing, where this is usually taken to mean testing hypothesized relationships between variables. And this is not because case studies are approached theory free. Rather, so it is claimed, case studies are research in the context and service of discovery, not justification: they are for the formation of evidence-based concepts, for the development of measurement structures, the places where types are defined and kinds isolated, where phenomena might be revealed and theory developed. (2012, 671)

According to Morgan (2012), case studies play an important role in revealing phenomena and accounting for phenomena, a capacity also attributed to experiments in the social interactionist and ethnomethodologist's traditions as discussed earlier in the Chapter. Case studies can elicit accounts which have not yet been explained by deploying a particular configuration and mode of togetherness. When they reveal and account for phenomena, case studies can act as vehicles of discovery. Instances of such cases have been called 'revelatory' (Yin, cited in Morgan 2012) when either the case serves to identify a phenomenon previously inaccessible to scientific investigation, or when cases are moulded in a way that brings previously not established conceptual associations into being that provide a novel perspective on different empirical phenomena. The difficulty but also advantage of these types of casing designs is that it is sometimes challenging to draw lines between a particular object of study and its context because of 'the open-endedness of the research

questions and because the object emerges to be fully distinguished from its context only during the course of the research.' (Morgan 2012, 672).

In my thesis I adopt this general understanding of casing, as I configure the empirical and the conceptual by way of a particular mode of bounded togetherness in the form of two cases for the study of data frames. Configuring the empirical material as cases has the experimental effect of accounting for, re-presenting and revealing the framing of data in novel ways. The bounding of these materials, devices and techniques as pertaining to cases has been mainly informed by the simultaneous conceptual and empirical work undertaken throughout the research process. This allows the materials taken into account to express and reflect conceptually relevant aspects, and enables different lines of enquiry to unfold in relation to what a range of data types like search queries and prices are and do. Furthermore, it facilitates the use of the materials as a means of problematising different phenomena like *personalisation* in the case of the AOL leak, or *composition* in the case of the CPI leak. In other words, the objective behind deploying casing as a research strategy is to link the empirical materials with a range of different conceptual ideas and to use the latter as a means of producing meaningful and revelatory descriptions of the framing of data and the devices that contribute to such framings. I use such descriptions to simultaneously advance, redefine and problematise conceptual ideas further.

The work of drawing lines that can identify and account for social scientific objects of study involves identification, description and conceptualisation, and as such is a process more closely associated with discovery than justification. The advantage of approaching and using casing in this way has been that it enabled me to maintain the

open-endedness of the case until quite advanced stages of the research process, and to make a virtue of the elusiveness of the object of study. This allows the research to branch out into new areas and materials from the initial base – in this case, the leaks themselves - and to maintain a blurred and flexible boundary between object of study and context in order to develop exploratory accounts that render phenomena visible in novel ways.

It could be argued that repurposing data leaks by way of a casing tactic has entailed designing a form of 'provocative containment' (Lezaun et al. 2012). Lezaun et al. have suggested that 'provocative containment' is a technique for the production and display of social reality' (2012, 280). Provocation – an activity which encompasses many social scientific techniques of intervention including but not exclusive to experiments - constitutes the triggering of an effect that unleashes the surfacing of something not readily available, leading to the pronunciation and display of something new (Lezaun et al. 2012, 280). Containment on the other hand, should be read as the handling of the empirical and conceptual 'inside', a clearly demarcated space, in order to 'hold it stationary, or in a manageable scale and duration, to prevent it from overflowing' (2012, 280). Lezaun et al. further note that 'containment also has to do with authorship and responsibility: the act of containing raises the issue of the elliptic position of the "third person", that is of the effective but partially elusive presence of the scientist who controls and devises' (2012, 280).

Casing has many of the characteristics of provocative containment, a technique attributed to experimental research. Whilst repurposing leaks by casing them is to retain some of the benefits of experimental techniques - such as that of provocation -

it also presupposes a different form of control. Containment by casing of course entails some form of control; but this control is explicitly geared towards a type of isolation that, instead of reducing and deleting, thickens, expands and is suggestive of further relations.

Conclusion

In this Chapter I have developed the notion of leak as an inventive method in order to shed light on how the former can be taken as provoking empirical occasions, and be repurposed for social scientific research. I have considered how the repurposing of leaks understood as a type of inventive method can be achieved by engaging with aspects of two other distinctive types of inventive methods: that of the *experiment* and that of the *case*. I have suggested that leaks can be considered as a form of societal experiment. I have compared society's experiments (Morgan 2012) to laboratory experiments, and highlighted their differences. I have also shown how society's experiments have the capacity to render phenomena, and frames, visible and accountable; and how they can be used as techniques for the production and display of social reality in a demonstrative and explicit form (Lezaun et al. 2012). I have then justified why and in which ways the leaks on which this research project focuses on can be taken as types of society's experiments, and the reasons why the repurposing of these leaks for social research purposes is more beneficial if the leaks are redesigned as cases rather than as experiments. The last section of the Chapter has looked into the research activity of casing and has identified the capacity of this activity as a mode of revealing and accounting for phenomena or as a revelatory research design. I also have suggested that casing resembles many of the

characteristics of provocative containment as described in Lezaun et al. (2012), while employing a different form of control.

I will turn now to the two cases that repurpose two different data leaks. The first case investigates the framing of data as personal or impersonal by a range of devices. It focuses on analysing how different devices and techniques work to configure certain pieces of data like search keywords to be about persons. Firstly, it looks at the technique of *reidentification demonstrations*, a method for identifying persons and making them visible from anonymised databases. Secondly, it addresses the *sequenced organisation* of search keywords as a means of enacting a particular mode of being a person. Thirdly, it looks at *anonymisation* and *informed consent* as devices that aid in the economisation of data and its depersonalisation. The second case examines a range of techniques and devices that stabilise the framing of price in the measurement and commercialisation of consumer price indexes. Firstly, it looks at how procedures for the *identification of products* aid in the establishment of prices as pure, and how legal aids like *Secrecy Acts* contain the ontological formation of price. Secondly, it investigates how digital measuring methods like *scraping* and *imaging* are reconfiguring the framing of retail price for financial use.

Case Study I - The Making and Unmaking of Personal Data

Introduction

In 2008, a flowchart was distributed by the Information Commissioner Office (ICO) in the United Kingdom as part of its Anonymisation Code of Practice. Both the flowchart and the code of practice were published and circulated to organisations and individuals who handle and manipulate different kinds of information, to aid them in determining when information is to be considered *personal data.* The flowchart, and the Code of Practice it was embedded in, was issued under the Data Protection Act 1998,[38] which is the piece of legislation that regulates what count as personal data in the UK, and how it should be dealt with. The flowchart proposes that data controllers - that is, individuals or organisations managing different kinds of information - should ask the following questions when presented with data deemed to be personal:

> Can an individual be *identified* from the data? Does the data *relate to* the identifiable living individual? Is the data *obviously about* a particular living individual? Is the data *linked to* a particular individual so that it provides particular information about that individual? Is the data used, or is it to be used, to *inform or influence decisions* affecting an identifiable individual? Does the data have any *biographical significance* in relation to the individual? Does the data *focus or concentrate on* the individual as its central theme

[38] Prompted by the introduction of the European Data Protection Directive in 1995.

> rather than on some other person, or some object, transaction or event? Does the data impact or has the *potential to impact* on an individual, whether in a personal, family, business, or professional capacity?[39]

Each combinatory possibility offers a different outcome that determines whether, from a legal point of view, 'data' - opinions, transactions, interactions, movements, beliefs, sentiments, and conversations, to name a few - should be considered personal or not.

The first legal case that was decisive in configuring a legal interpretation of personal data in the United Kingdom was that of *Durant vs. Financial Services Authority (FSA)*. This case was paradigmatic in framing a definition of personal data for the legal system. The claimant Durant was in dispute with Barclays Bank, and made a complaint to the Financial Services Authority (FSA). This led to a confidential enquiry by the FSA into the bank's proceedings in relation to Durant. The claimant, who had already lost various law suits against the bank, requested access to all records held by the FSA 'which mentioned his name or in were in any way "related to" him, on the grounds that they were "personal data" of which he was the subject and to which, by ss7(1) and 8(2) of the 1998 Act, he thus had rights of access' (Edwards 2004, 344). The legal case focused on defining 'how widely the phrase "relate to" in s1(1) should be interpreted' (Edwards 2004, 344). A wide or narrow definition of the phrase 'relate to' would ultimately determine what information the claimant had the right to see. The English Court of Appeal opted for a narrow

[39] Information Comissioner Office, Draft Anonymisation Code of Practice, for Consultation 31 May to 23rd August 2010.

definition of personal data, based on one of two definitions taken from the Shorter Oxford Dictionary for the phrase 'relate to' (Edwards 2001, 343). The narrower definition stated that the phrase 'relate to' should be understood as '"having reference to, concern'" (Edwards 2004, 343), whereas the wider definition should be understood as '"having some connection with, be connected to"' (Edwards 2004, 323). It was argued that the narrower definition was more appropriate to serve the legal purpose of the 1998 Act because it entailed giving Durant 'information about himself' only, and not information specific to any other things or persons per se, like documents or third party individuals. From then onwards from a legal perspective, 'whether any particular information amounted to "personal data" would in general depend on where it fell in a *"continuum of relevance or proximity"* to the data subject' (Edwards 2004, 344; my italics).

I use these juridical examples as an introduction to this case because they clearly illustrate how the configuration of data as personal is dependent on different framings and framing devices operating within a range of fields in society at large. The flowchart can be thought of as a framing device that, when applied to a particular case, can determine whether a piece of data is deemed to relate to or be about a person. The legal dispute however brings into focus how the application of dictionary definitions was instrumental in defining as impersonal the relation between a physical person and some documents containing that person's name. The continuum of relevance and proximity between persons and data is articulated by framing devices that operate to extend and collapse this relevance and proximity. This case explores how the introduction of relatively new digital technologies like web-based search engines re-define the relation between data and persons. It also investigates how the

relation between data and persons has been historically configured by well-established devices like informed consent and anonymisation, traditionally deployed as part of ethical research protocols in a range of scientific disciplines.

The case study repurposes a search keyword data leak, the events and materials that this leak unleashed, and the devices and techniques for the framing of data that the leak either afforded or rendered visible for analysis. These are 'cased' as a way of studying the making and unmaking of personal data via the workings of techniques and devices, looking at how data becomes contextualised as being about persons on the one hand, and detached and disentangled from persons on the other. The case study suggests that data becomes personalised or depersonalised depending on the workings of different framing techniques and devices. The overflow of data and its subsequent reconfiguration as a case study makes it possible to observe how framing devices of data personalisation and depersonalisation work, and how they are being deployed as part of different practices, genres and protocols. The first technique that I want to engage with is that of a *reidentification demonstration*. Reidentification demonstrations are statistical techniques deployed for identifying and invoking persons out of anonymised databases. The case study also discusses different techniques being used in the production of data art. One of these techniques is the *sequential reordering* of data, that is, the repurposing of data found organized in a serial order in a database. The second data art technique is the articulation of a *nameless biography*, that is, a biography whose function is independent of the use of a personal name. Finally, the case study also looks at the devices of *informed consent* and *anonymisation* as part of research ethics protocols that became visible as part of the data leak. Informed consent is a procedure which, by securing a particular form of

informational exchange, provokes research subjects to arrive at a decision about their participation in different forms of data extraction and manipulation. Anonymisation is a device designed and deployed to delete the link between persons and data.

The case is comprised of two Chapters. The first Chapter – entitled *The Technical Redoing and the Aesthetic Reordering of Search Keywords* - looks at the emergence of search keywords as personal data, and analyses the devices and techniques which enable this articulation to take place. It describes how search engines have configured search keywords as an expression of subjective intentionality. The Chapter then moves on to analyse the techniques of reidentification demonstration, sequential reordering and nameless biography as described above. I argue that the public reidentification demonstration of a woman via the re-enactment of her search query history constitutes a framing technique for domesticating search keywords as personal. I also claim that reidentification demonstrations can be conceived of as framing techniques that articulate 'merographic connections' (Strathern 1992), and as mechanisms for the 'technical redoing' (Goffman 1974) of data. The Chapter then proceeds to analyse a piece of data art that was also created with the aid of materials that emerged from the leak. I analyse this artwork as a frame that enables search keywords to become configured as personal, but more specifically as a frame that facilitates a nameless biographical account to emerge. I suggest that the artwork challenges the historical function that personal names have had in configuring different data arrangements as personal. This Chapter concludes by offering some reflections on whether *traceability* (Latour 1999, 2010) - the framing of data to refer back to a point of origin - can be understood as a form of technical redoing (Goffman

1974) and if so, whether this means that processes of inscribing or making data should be understood as irreversible.

In the second Chapter, *Ethical Devices and the Economisation of Personal Data,* the case study focuses on how the devices of informed consent and anonymisation enable a particular economy of data to unfold by disentangling and reframing the relation between data and persons in specific ways. I firstly explore the device of informed consent and claim that it aids in the enactment of a particular type of research subject equipped with autonomy, self-determination and entrenched in frames of privacy. I also suggest that the device of informed consent enables a transfer of data which would otherwise be deemed illegitimate, and can thus be understood as a 'surrogate property contract' (Waldby and Mitchell 2006). Secondly, the Chapter looks at the device of anonymisation as a data economisation device. It examines how the economisation of data requires the framing of data as separate from the person to which it was linked, a separation which is enabled by anonymisation. The Chapter also discuses how the anonymisation of data can be thought of as a device that suppresses the author function, thereby undoing data's capacities to be an object that can be appropriated and owned by data subjects. In the conclusion, the Chapter offers suggestions at what should be taken into account in the design of research ethics for newly emerging datascapes and analytical techniques.

Chapter Three

The Technical Redoing and the Aesthetic Reordering of Search Keywords

Introduction

What is personal data? Inscriptions that today are taken to be biographical or about persons did not qualify as such before the emergence of a new range of devices and technologies that deliberately or indirectly record human agency as part of contemporary everyday life. The gathering of what is currently taken for granted as 'personal data' is becoming more pervasive, as interactions with new digital technologies and devices become embedded in daily life routines. Objective and subjective statements, behaviours and characteristics, geographical locations, biological properties, physiological patterns, living traits, actions and opinions - all are now allegedly being inscribed, recorded and stored at higher rates than ever before. At the same time that personal data appears to proliferate more extensively, technologies are also being designed to regulate its circulation[40] and to obfuscate its epistemic status[41] as a way of preserving particular versions of personhood – predominantly as private. Different examples of what constitutes personal data have flourished in response: 'the tall elderly man with a dachshund who lives at Number 15 and drives a Porsche Cayenne' (Information Commissioner Office Personal Data

[40] See for example Privacy Enhancing Technologies (PET) in Acquisti et al. (2008).

[41] See for example TrackMeNot (TMN) at http://cs.nyu.edu/trackmenot/, last accessed 16th November 2012. Also see Nissembaum H. and Brunton F. (2015).

Guidance, 2007). The EU Data Protection Directive also declared on the concept of personal data that

> From the point of view of the nature of the information, the concept of personal data includes any sort of statements about a person. It covers 'objective' information, such as the presence of a certain substance in one's blood. It also includes 'subjective' information, opinions or assessments.[42]

In this Chapter, I would like to trace the emergence of search keywords as personal data and as data about persons. Search keywords are a relatively new data type that has emerged with the incorporation of search interfaces into different sorts of digital networked devices, including the Worldwide Web, operating systems, software, mobile applications and web-based services. Most prominently, web-based search engines have been prevalent in configuring the relation between persons and search keywords, particularly since they started to personalise the results that they offer to users, from the early 2000s onwards. However, at their inception, the search results that web-based search engines offered were conceived of as depersonalised. Engines offered search results that were considered relevant at the aggregate level in terms of how the index reorganised itself with regards to each particular search query and used the method of link analysis to construct a ranking made visible as a query list.[43] However, this search data was still considered to be mostly irrelevant for the specific needs and subjective interests of the individually situated users that had

---

[42] See Article 29 Data protection Working Party, Opinion 4/2007 on the Concept of Personal Data, EU Data Protection Directive 1995.

[43] See for example Pasquinelli (2009).

operationalised the search (Stalder and Mayer 2009). Martin Feuz et al. stated in this respect that

> As the range of information, context and users of Internet grew, the relationship between the search query, search interest and user became more tenuous. Not all users were seeking the same information, even if they used the same query term, and even individual users did not connect the same search interest with a particular query at all times and locations. Thus, the quality of search results, at least potentially, decreased. Search engines started to respond to this problem by trying to personalise searching, promising to deliver more relevant results to the user, whose query was being considered in the context of his/her search history and other data complied into a personal profile. (2011).

Due to proprietary restrictions it has been difficult so far to publicly assess the types and extent of personalising functions enabled by commercial search engines. Some research in the area has been geared towards describing all the publicly visible mechanisms through which personalisation is accomplished by search engines (Stalder and Mayer 2009); others have deployed digital experiments to test a given search engine's personalising capacity and effects (Feuz et al. 2011); and different streams of critique have emerged from historical and cultural accounts of web searching (see for example König and Rash 2014). These are by no means thorough examinations of the personalising effects of search, but they offer a starting point to think about what personalisation might imply for the relation that search engines establish between users on one hand, and search queries on the other.

The technologies that underpin web-based search have been described as a 'database of intentions' or as 'a massive clickstream database of desires, needs, wants, and preferences that can be discovered, subpoenaed, archived, tracked, and exploited for all sorts of ends' (Battelle, cited in Jarrett 2014, 14). Kylie Jarrett argued for example in relation to Google that

> This database is a collation of each search term entered by a user and various other kinds of data, such as the geographic location of the computer's IP address, the time spent on particular sites, and user preferences gleaned from profiles on various related sites such as YouTube, GoogleBooks, Gmail, or Google+. These data are re-integrated into the Google infrastructure in various forms: as market demographic information, as personalization algorithms that refine search parameters, as mechanisms for algorithms to 'learn' natural language use, or as evidence of collective wisdom in 'trending' statistics, to name but a few instances of re-purposing. (2014, 17)

According to Feuz et al. (2011), the personalisation of search was firstly enabled by the compilation of a personal profile based on the assumption that each particular search query should be attributed to a singular and unique user interest or need. Personalisation also demanded that the context in which search queries occurred was provided by a range of other data types, including: background long-term preferences; search query history; browsing history; semantic scope of the information requested; user actions not related to the search in itself and so on. So while users' interests and needs were initially conceived of as being unaffected by

their interaction with the retrieval or search arrangement they participated in, the advent of personalised search results and the ubiquitous presence of search engines meant that search engines started playing an important role in determining what users ultimately searched for and how.

A key aspect of the development of personalising functions as part of web-based search is their pre-emptive capacities. This has been made possible by a number of analytic operations, including: the creation of personal profiles; the semantic determination of search keywords based on history search and other data models; the identification of geo-demographical coordinates; the deployment of algorithms that learn naturally-occurring language use and refine search parameters; and auto-complete and auto-suggest mechanisms that display personalised search paths. According to Jarrett (2014), the model of relevance used by web search engines in their deployment of different personalising technologies, search functions and data models 'draws on data of each user's past behaviour along with aggregated generic data to propose a future model' (23). She further argues that

> This is then mobilized to propose personalized and yet generic search results and search suggestions shaped for that user. Thus, the intentions ascribed to each individual, constituted by an amalgam of that user's search practices and the intentional logic ascribed to users as a mass, are deployed to guess the desires and orientations of each user and preemptively structure the search experience. (Jarret 2014, 23)

Jarrett suggests that through the amalgamation of these different technologies and

techniques, a feedback loop emerges in which presumptions, based on the search engine's configuration of user's activities (including the typing in of keywords), go on to inform the user's experience of, although not necessarily engagement with, search (2014, 24). In other words, search engines configure, format and anticipate the limits, contents and possibilities of interests and needs expressed in search keywords. It could be argued that search engines configure the relation between persons and search keywords in such a way that the contextualisation of the latter via other data types (but also past search keywords themselves) enables a pre-emptive determination and shaping of what the search intentions of a person are and can be. The active contextualisation of search queries as part of the operations of search engines therefore configures keywords as constitutive of uniquely subjectified persons with individual search intentions and needs.

The personalisation of search is one way in which search keywords become related and entangled with persons. The personalisation of search as it is pursued by web-based engines can be understood as a relatively new technical mode of articulating particular versions of persons – and their search intentionality - via the practice of typing search keywords into a search engine. That personalisation is being undertaken via search practices should be considered as nothing less than problematic, according to critique advanced by a range of scholars (König and Rash 2014; Feuz et al. 2011; Stalder and Mayer 2009). Felix Stalder and Christine Mayer (2009) suggest that since the development of digital web-based information retrieval indexes based on users' history and context, search engines have become prominent in operationalising subjective worlds. These worlds are determined by an opaque (to the user) reconfiguration of the individual's search histories and contexts that delimit what

information is suitable and relevant for their subjective interest at a given point in time based on proprietary knowledge. This constitutes a system that, so Stalder and Mayer (2009) claim, ultimately functions through a profound loss of subject autonomy and a lack of transparency over how personalising functions of search ultimately operate.

Here, I have briefly discussed how search keywords are configured through search engines' operations as an expression of subjective informational intentionality. That web-based search engines operate to produce this relation does not however entail that the material expression of search – in this case a string of digital typed-in text known as a 'keyword' – has been historically associated with persons. In suggesting this, I am not proposing that before the emergence of web-based search there were no necessarily connections between the practices of searching and persons; but that searching as an act of retrieving information has become an increasingly important mode of making up and equipping persons, and that this in turn is unleashing new potential modalities of experiencing particular ways of being a person (Hacking 1986). The emergence of keywords understood as the material expression of search on one hand and persons on the other is, I furthermore argue, not only being configured by search engines but is also being negotiated in the execution of wider social and public experiments with data. In the following sections I show how the domestication or incorporation of new technical elements in society, in this case of search keywords as part of persons, has been undertaken via two different techniques: a reidentification demonstration undertaken by two journalists; and the reorganisation of search keywords as part of an art piece.

Reidentification Demonstrations

Reidentification demonstrations[44] consist in the deployment of different techniques – including the design of algorithms and the amalgamation of different databases – in order to trace, identify and name persons from databases containing purportedly anonymised data. The phenomenon originated in the activities undertaken by Professor Latanaya Sweeney, now director of the Data Privacy Lab at Harvard University. In the mid 1990s, when Professor Sweeney was an MIT graduate student, the Massachusetts Group Insurance Commission (GIC) responsible for buying health insurance for state employees collected and released anonymous patient-specific data containing nearly one hundred attributes per 135,000 visits to hospital. As the data was anonymous, GIC deemed it appropiate to distribute its database for research purposes and also sold copies to industry-related organisations (Sweeney 2002). Reassurance regarding privacy concerns was also given by Massachusetts Governor Bill Weld, who guaranteed that, due to the anonymisation procedures conducted, the risk of personal public reidentification via the published data had been made inexistent.

Unfortunately, Weld later collapsed at a local public event and was admitted to hospital, an occasion that Sweeney quickly saw the potential in. She purchased 'the voter registration list for Cambridge Massachusetts' for twenty dollars and 'received the information on two diskettes' (Sweeney 2002, 2). The data in the diskettes

[44] See http://blogs.law.harvard.edu/billofhealth/category/re-identification-symposium/, last accessed 5th September 2015.

contained the name, address, ZIP code, birth date and gender of every voter in Massachusetts. Sweeney combined this data with GIC records in order to identify Governor Weld with his hospital visit. She found that only six people in the city of Cambridge shared Weld's birth date, of those only three were also men and only one - Weld himself - lived in the area defined by his postcode. In a rather 'theatrical flourish' (Ohm 2009, 18), Sweeney posted the records – which included diagnoses and prescriptions - to Weld personally.

Since then a range of different reidentification demonstrations have taken place, conducted by academics, citizens and a variety of other actors, including most prominently the case of Netflix Prize Data release,[45] the so-called AOL data leak,[46] and also Sweeney's most recent engagement with the Personal Genome Project[47] database, to name a few. Reidentification demonstrations are deliberately staged and publicised, among others as examples of how privacy can be breeched if data is not anonymised properly. Computer scientists in particular have undertaken reidentification demonstrations as a means of disrupting the privacy policy landscape, by showing that anonymisation techniques like the ones promoted by different laws in Europe[48] and America[49] are inadequate as a means of guaranteeing the protection of privacy. However, I would like to argue that reidentification demonstrations can also be conceived of as a form of public experiment deployed for the domestication

[45] See Narayanan and Shmatikov (2008).

[46] See below for a detailed description.

[47] See Sweeney et al. (2013).

[48] See European Data Protection Directive 1995.

[49] In the United States, federal privacy statutes carve out exceptions for those who anonymise.

of data types. Techniques like reidentification demonstrations blur the boundaries between the realms of techno-science and society (Michael 2006), and are deployed to renegotiate the meaning and capacities attributed to entities like data units. In other words, when performed in public, *reidentification demonstrations can be thought of as techniques for the societal domestication of data, by framing but also contesting what units of different database sets are ultimately 'about'*. This is arguably a process of both re-domestication and de-domestication, as it also renders the familiar routines of data extraction and manipulation strange. Let me here elaborate on this point further by engaging with the case of AOL's data leak.

On the 3rd of August 2006, Abdur Chowdhury, then head of American Online (AOL) Research, sent an email to the Special Interest Group on Information Retrieval (SIGIR)[50] mailing list (among other addresses):

> AOL is embarking on a new direction for its business - making its content and products freely available to all consumers. To support those goals, AOL is also embracing the vision of an open research community, which is creating opportunities for researchers in academia and industry alike.
>
> We are introducing AOL Research to everyone, with the goal of facilitating closer collaboration between AOL and anyone with a desire to work on interesting problems. To get started, we invite you to visit us at http://research.aol.com, where you will find:

[50] See http://sigir.org, last accessed 9th August 2015.

- 20,000 hand labelled, classified queries

- 3.5 million web question/answer queries (who, what, where, when, etc.)

- Query streams for 500,000 users over 3 months (20 million queries)

- Query arrival rates for queuing analysis

- 2 million queries against US Government domains

Also, please feel free to provide feedback on the site, datasets you'd like to see in the future, and any other comments about our vision.

Thanks,

Abdur Chowdhury[51]

The data was made accessible alongside an accompanying *read me* file (see Figure 1.0 in the Appendix), which further explained the nature and characteristics of the data set released. AOL's transfer of data to academia was seen by many as a positive step towards making data - which until then had been generally confined to

[51] AOL's Head of Research Email, http://gregsadetsky.com/aol-data/, last accessed 2nd January 2013.

researchers working in web-search companies such as AOL - open and accessible.[52] The data, which was intended to be used for research purposes only was later reassembled and redeployed in alternative ways by a number of diverse actors, including journalists, artists, performers[53] and bloggers.[54] What came to be known by the public and those involved as the 'AOL Data Leak' consisted in the proliferation of experimental and alternative data arrangements which contested the framing of search keywords as simple, effortless expressions of a priori naturally occurring subjective informational intentions and needs.

Despite the fact that the AOL data had been anonymised, Michael Barbaro and Tom Zeller, two journalists at the New York Times, used AOL's files to pinpoint different clues to identity in search keywords by combining the leaked database with phonebook records. By examining user number 4417749's queries, they searched for and found a person. The journalists claimed that

---

[52] Before the release of AOL's data, academics in the information retrieval research field in America claimed to have worked for nearly ten years with two sets of corporate search data deemed to be no longer relevant by 2006: one from Excite and one from Altavista. This is one of the reasons why the release of relatively fresh and raw search query data was initially welcomed in the field of information retrieval or search research at the time, particularly as it provided a way of understanding how people asked questions for information retrieval purposes at that particular point in time (as opposed to working with either old or synthetically generated data). An article on The New York Times on August the 23rd of August illustrates these positions succinctly: see http://www.nytimes.com/2006/08/23/technology/, last accessed 9th August 2015.

[53] A theatrical play produced by a Philadelphia theatre company entitled "User 927", directed by Katherine Clark Grey. The play ran from June 11 through to June 22 2007 at the Studio at St. Stephen's Theatre in Philadelphia.

[54] See for example http://search-id.com and http://www.aolstalker.com, last accessed August 9th 2015.

> Number 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything".
>
> And search by search, click by click, the identity of AOL's user number 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga", several people with the last name Arnold and "homes sold in shadow lake subdivision Gwinnett county Georgia."
>
> It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches", she said, after a reporter read part of the list to her.[55]

Figure 1.1 and Figure 1.2 in the Appendix show Thelma Arnold and her searches. Thelma Arnold claimed she 'was shocked to hear' that AOL had saved and published three months' worth of her search queries. The New York Times reported that Thelma Arnold referred to such keywords as her 'whole personal life'. The journalists continued:

---

[55] Barbaro and Zeller (2006, August 9), A Face Is Exposed for AOL Searcher No. 4417749. *The New York Times*. Retrieved from http://www.nytimes.com, last accessed 9th August 2015.

> In the privacy of her four-bedroom home, Ms. Arnold searched for the answers to scores of life's questions, big and small. How could she buy "school supplies for Iraq children"? What is the "safest place to live"? What is "the best season to visit Italy"? Her searches are a catalog of intentions, curiosity, anxieties and quotidian questions. There was the day in May, for example, when she typed in "termites," then "tea for good health" then "mature living," all within a few hours.
>
> Her queries mirror millions of those captured in AOL's database, which reveal the concerns of expectant mothers, cancer patients, college students and music lovers. User No. 2178 searches for "foods to avoid when breast feeding." No. 3482401 seeks guidance on "calorie counting." No. 3483689 searches for the songs "Time After Time" and "Wind Beneath My Wings."
>
> At times, the searches appear to betray intimate emotions and personal dilemmas. No. 3505202 asks about "depression and medical leave." No. 7268042 types "fear that spouse contemplating cheating."[56]

It has been suggested that as soon as AOL's data was made public both 'AOL and the rest of the world learned that search engines queries are windows to the soul' (Ohm 2009, 16). I argue, however, that search queries were not in fact inherently personal to start with - they *became* articulated as personal. In short, there is no essential

---

[56] Barbaro and Zeller (2006, August 9), A Face Is Exposed for AOL Searcher No. 4417749. *The New York Times*. Retrieved from http://www.nytimes.com, last accessed 9th August 2015.

distinction in principle between personal and impersonal data. Search keywords are not the outcome of an unmediated expression of subjectivity; rather, certain pieces of data are made to be personal, or about persons, by way of framings, techniques and devices.

Reidentification demonstrations can be understood as one such technique. It was deployed in the case of the AOL data leak to publicly augment what was deemed personal, and to trigger a reaction by which a concatenation of relatively meaningful but sometimes misspelt words typed into an interface became an expression of intimacy and privacy. In the retrofitted and orchestrated encounter between Thelma Arnold's person and the search queries she allegedly produced, search keywords emerged as an inscription of interiority and as an expression of personal dimensionality (Berlant 2000). As Ian Hacking (1986) argues, possible ways of being a person can come into being and disappear, and as such there are degrees of possibility for being a person in particular places and times. Reidentification demonstrations like the one here described open up the possibility not only of relating the realm of interiority with that of digital search, but also for individuals – here Thelma Arnold - to experience themselves and be experienced - as being particular persons inserted in the particular context of their privacy through search.

There is however an important element that needs to be taken into account if we are to consider reidentification demonstrations as a means of making data personal. Reidentification demonstrations - like the one that configured search keywords as related to Thelma Arnold - are made in public and for a public. From the perspective of STS, the function of events like public experiments is to domesticate – to

naturalise and adapt - techno-science into public and social life. They are seen as ceremonial occasions for the introduction of new entities that 'reconfigure the wider social-material relations among which the (new) object is to be accommodated' (Marres 2009, 119). In this sense, AOL's reidentification demonstration served as a way of introducing a relatively new data type like search keywords into society. This event rendered and reorganised the meaning and experience of this data type as personal, and re-inscribed what counts as personal and private in society at large. Genevieve Bell (2015) has recently suggested that data 'bubbles up to the surface in all sorts of places' and that 'it finds its space in a cartography of social imagination' (p. 10). More specifically, the social configuration (and imagination) of data is being articulated through the techniques being deployed for its domestication.

However, if we are to think about the reidentification demonstration described above as a form of public experiment with data, we also have to think about how such experiments are achieved. What are the elements and techniques that form part of the experiment as a means of testing, displaying and reconfiguring data in new ways? I here would like to argue that whilst there are no essential distinctions between what counts as personal or impersonal data, public experiments designed to relate certain pieces of data with particular types of persons establish what Marilyn Strathern (1992) calls 'merographic connections'. Sarah Franklin (2003) argues that 'merographic connections' are introduced in Strathern's discussion of 'biological and social facts in English kinship thinking as a means of distinguishing a specific kind of cultural borrowing involved in the way ideas travel, connect, disconnect and contain one another' (2003, 66). Strathern (1992) uses the distinction between 'wholes' and 'parts' (or multiple different 'parts') 'to explore the ways in which 'parts' overlap in

the production of ideas about relatedness' (Franklin 2003, 67). As an example, Franklin claims that kinship and new genetics connect the realms of the natural and the social 'merographically' because within the idea of a kinship marker such as DNA there is a 'co-mingling of parts that belong to different wholes' (2003, 67).

As part of the reidentification demonstration described above, search keywords were reclaimed and re-described as parts and expressions of personhood. A merographic connection, according to Strathern (1992), occurs not when an entity is substituted by another as a version of itself, but when such substitution connects the entity to a wholly different domain or set of phenomena. Strathern argues that 'the very desire to put facts "into their context" is a merographic move. The context, by virtue of not being equivalent with the thing put into it, will "illuminate" the thing from a particular angle (display one of its parts)' (1992, 73). While data, instantiated as a typed-in search keyword, can be understood as part of a person as a whole, it can also be conceived of as a part of other wholes - as part of a series within a dataset, for example, and also as part of a person's interiority or subjectivity. When search queries are reconfigured as part of Thelma Arnold in a way that also serves to elicit her private context, a co-mingling of parts that can belong to multiple different wholes occurs.

Furthermore, I would like to add that apart from exhibiting 'merographic connections', the experimental apparatus of public reidentification demonstrations also make use of techniques of reframing that Goffman (1974) conceptualised as 'technical redoings'. Technical redoings were one of the frame rekeyings that Goffman typified as prominent social practices at the time he was writing. More

precisely, he described them as 'strips of what could have been ordinary activity that are performed out of their usual context, for utilitarian purposes openly different from those of the original performance' (1974, 59). He distinguished four different varieties of technical redoings: simulations; demonstrations; documentations; and experiments (1974).

Goffman considered *simulations* or trial sessions as instances in which 'events become decoupled from their usual embedment in consequentiality' (1974, 59). *Demonstrations* were taken as 'performances of a task-like activity out of its usual functional context in order to allow someone who is not the performer to obtain a close picture of the doing of the activity' (Goffman 1974, 66). Goffman noted with regards to *documentations* that 'in society there is considerable (and growing) use of replicative records of events, that is replays of a recording of a strip of actual activity for the purpose of establishing as fact, as having occurred, something that happened in the past' (1974, 68). Finally, Goffman proposed that *experiments* should be seen as attempts at maintaining natural conditions when natural reasons do not exist for such a performance. He specified that in order for an experiment to be considered as a rekeying, it should be assumed that 'the participants in the activity – experimenters, subjects, audience, etc. – all share the same understanding of what is happening when it is happening, namely, an experiment of a particular kind' (1964, 73).

These different types of technical redoings have several characteristics in common. Firstly, they are all conceived of as events geared towards performing an original strip of activity through staging a different context for it, or putting an object to work for a different purpose than that originally intended for it. Secondly, technical

redoings are not presented as attempts to copy or replicate the original event or activity, but rather are understood as patterned on them to achieve a different effect, provoking in turn a different organisation of meaning and experience. For Goffman (1974), this patterned decoupling entails introducing a different frame that could differentiate itself from the original activity on which it was patterned. This is achieved by establishing numerous cues that would indicate when a transformation began and when it ended; these cues were understood as temporal and spatial brackets or devices which indicated 'everywhere within which and nowhere outside of which the keying applies on that occasion' (Goffman 1974, 45). In other words, technical redoings require different degrees of designed de-contextualisation and re-contextualisation; in the case of technical artefacts, like data units or search keywords; they therefore might be thought of more generally as 'experiments in contexting' (Asdal and Moser 2012).

Data Sequencing

As I have described in the previous section, while AOL's data release was intended for research purposes only, the dataset overflowed and was reframed and rendered public by a range of actors in different ways. In the following section, I turn to focus on a data art work based on the search keywords that were leaked. I look at how the reordering and sequencing of search keywords in this artwork produced a nameless biographical account, and how this account again contributed to publicly reframe search keywords as pertaining to persons and, what is more, as part of particular biographical arrangements. According to Victoria Vesna (2007), artists have long

acknowledged the conceptual and aesthetic possibilities of databases and taken these as 'ready-made commentaries on our contemporary social and political lives' (2007, 6). However, Lev Manovich argues (2001) that many new media objects like databases do not necessarily tell stories, as they are not configured to deliver a beginning or an end and they do not follow any thematic or formal development. The elements within many new media objects, including the database, 'are not organized as sequences but are instead collections: serial catalogues of individual items, with every item possessing the same relevance and saliency as any other' (Manovich 2001, 218). Manovich maintains that while cinema privileged narrative as the key form of cultural expression of modern age, the computer age introduced its correlate: the database. He continues by suggesting that

> As a cultural form, the database represents the world as a list of items and it refuses to order this list. In contrast, a narrative creates a cause-and-effect trajectory of seemingly unordered items (events). Therefore, database and narrative are natural enemies. Competing for the same territory of human culture, each claims an exclusive right to make meaning out of the world. (2001, 225)

One could argue however that even if databases do not tell stories, the seriality afforded by the framing of data units as part of databases is still able to 'pose a set of pervasive and prominent questions about continuity versus discontinuity, the play of difference through standardized objects, and the sequential display versus the array that could be seen at a glance' (Hopwood et al. 2010, 252). Further, while at first glance there is nothing in the serial logic of the database as media technology that

would foster versions of narrative biographies, this does not entail that biographical narratives cannot travel out of databases even when the function of personal names - so important for biographical renderings which bring together a range of traits and events as pertaining to a person - become suspended through anonymisation techniques. For example, a number of records within a database can be brought together in a particular trajectory, they can be repurposed in a particular order and alternative intervals, or flows, can be established among those elements. The reorganisation of a database and its records can bring about particular and alternative ways of doing biographies and invoking persons. By again following the trail of materials that emerged as part of AOL's leak, I here show how the reordering of search queries enabled partial biographical accounts to emerge, and how such reordering also permitted an experimental exploration of what search queries are ultimately about. I will now describe the art piece in question in order to elaborate on this further.

By using the leaked AOL's database, artists Lernert Engelberts and Sander Plug unfold AOL's User Number 711391 into a person over a series of thirteen short documentary pieces entitled *I Love Alaska – The Heartbreaking Search History of AOL's User 711391.*[57] The documentaries[58] were produced according to a specific format that combined thirteen episodic documentaries of no longer than seven

[57] Data art is an expanding field of practice. Other art pieces which use so-called personal data as materials include: We Feel Fine (Jonathan Harris and Sep Kamvar); I Want You, You Want Me (Jonathan Harris and Sep Kamvar); and The Movable Type (Ben Rubin and Mark Hansen).

[58] http://www.minimovies.org/documentaires/view/ilovealaska/, last accessed 21st August 2015.

minutes each, with the intention of developing a self-contained story. The *I Love Alaska* documentaries depict a barren, isolated landscape whilst a female narrator reads, in chronological order, a range of selected search queries made by an anonymous AOL user uniquely identified by the number 711391 as part of the leaked database.[59] The artists introduce their piece as follows:

> August 4, 2006, the personal search queries of 650,000 AOL (America Online) users accidentally ended up on the Internet, for all to see. These search queries were entered in AOL's search engine over a three-month period. After three days AOL realized their blunder and removed the data from their site, but the sensitive private data had already leaked to several other sites.
>
> I love Alaska tells the story of one of those AOL users. We get to know a religious middle-aged woman from Houston, Texas, who spends her days at home behind her TV and computer. Her unique style of phrasing combined with her putting her ideas, convictions and obsessions into AOL's search engine, turn her personal story into a disconcerting novel of sorts.[60]

[59] See a sample list of queries for user 711391 as organised by AOL's database in the Appendix (Figure 1.3).

[60] See http://www.lernert.nl/index.php?/projects/i-love-alaska/, last accessed 15th March 2015.

The episodes are organised chronologically but importantly each of them is also assigned a theme.[61] The narrative of each episode is taken from search keywords that, once arranged in a particular way, portray a person's profile and life events as they unfolded over the three-month period covered by the search data made accessible by AOL. For example, episode 7, entitled *user #711391 prepares for an upcoming important event,* is composed of the following search keywords read out in the following order, against a background of snowy images of nondescript countryside:

Continental Airlines

Why do men have food fetishes

Are you not supposed to go in the sun when you take antibiotics

How can a woman charm a man and make him laugh

Drug use was very common in the 1970

Did people get high and go to the rocky horror picture show

Can touching fingernails with fungus cause your fingernails to get fungus

[61] Episode 1 – March 1-7 2006: introducing user #711391 and her unique way of searching the Internet; Episode 2 – March 9-14 006: user #711391 shows an excessive preoccupation with her health and sleeping patterns; Episode 3 – March 15-22 2006: a latent same sex interest and continued hypochondria; Episode 4 – March 23-29 2006: user #711391 can't seem to remove stubborn make up; Episode 5 – March 30-April 5 2006: user #711391 questions the deeper meaning of online relationships; Episode 6 – April 6-12 2006: user #711391 investigates hair products and the art of interior decoration; Episode 7 – April 13-19 2006: user #711391 prepares for an upcoming important event; Episode 8 – April 20-26 2006: user #711391 does everything in her power to make a first good impression; Episode 9 – April 27-May 3 2006: user #711391 dusts off an old outfit for an eagerly awaited meeting with an online friend; Episode 10 – May 4-10 2006: user #711391 fundamentally regrets her infidelity; Episode 11 – May 11-16 2006: user #711391 tries to overcome her compulsion and asks god for forgiveness; Episode 12 – May 18-24 2006: user #711391 confronts her demons; Episode 13 – May 25-31 2006: user #71139 attempts to reconnect with her husband.

Don’t cut your hair before a big event

Cooking channel

What is prosciutto

What is capers

What is the gift of hospitality

Is Nicole Ritchie adopted

Can not drinking enough water cause a bladder infection

Crystal Bernard gay rumors

Keeping busy can help your nerves

Should you plan sex before meeting a cyberlover

Hotels in San Antonio, Texas

Brooke Burke seems like such a bitch

Brooke Shields has baby number two

Husband does not think is a good idea for me to meet my online friends

The narratives of the documentaries, however, are not only organised by the sequencing of different search phrases read in a neutral fashion, and naturalistic imagery. Intervals are introduced by the artists to make the sequencing of the searches explicit, and to emphasise the types of relations or connections established between individual queries. In episode 1 for example, the artists draw attention to the fact that they are introducing user number 711391 and her ‘unique way of searching the internet’. The selection of queries in this episode therefore appears to be thematically random, but elements are grouped together as they seem particular in the

way queries are phrased.[62] In episode 2, the artists suggest that user number 711391's queries show an excessive preoccupation with her health and sleeping patterns, although her queries are not all health- or sleep-related (nor do they necessarily contain either of these two words). The searches are also separated by a time interval that distinguishes the ending of one unit of search and the beginning of another one as a way of maintaining them as discrete entities.

The artists saw the documentaries as a way of making a personality visible, and they describe the woman who emerged from the documentaries as follows:

> Over a period of three months, a portrait of a woman emerges who is diligently searching for like-minded souls. The list of her search queries read aloud by a voice-over reads like a revealing character study of a somewhat obese middle-aged lady in her menopause, who is looking for a way to rejuvenate her sex life. In the end, when she cheats on her husband with a man she met online, her life seems to crumble around her. She regrets her deceit, admits to her Internet addiction and dreams of a new life in Alaska.[63]

In *I Love Alaska*, database seriality and cinematic sequencing are merged into a new form. In order to enact a person through data, the documentaries traverse and navigate the database, reorganising its records to produce a biographical narrative, without having to rely on the customary use of personal names. Such a method not

[62] Queries are phrased as if the user were interrogating another person in a naturally occurring situation.

[63] See http://www.lernert.nl/index.php?/projects/i-love-alaska/, last accessed 15th March 2015.

only produces a peculiar nameless biographical narrative; it also, I suggest, contributes to the reframing of search keywords themselves in a similar way to the framing that was accomplished by the reidentification demonstration described in the previous section of the Chapter. It was claimed by art commentators, for example, that although it would be unfair to presume that we have a coherent picture of user #711391 through *I Love Alaska's* biographical rendering, nevertheless 'the bottomless solitude of being trapped in one's own skin' is conjured up in each of the search keywords made public. The documentaries reframe a search history as 'an incidental archive of the self', where keywords 'now provide access into the fabled perpetual process of interior life: searching'.[64]

By only displaying the unique identifying number that had been assigned by AOL in order to make the data anonymous, the art piece demonstrates that by reordering database items, narrated biographical narrations can be produced without having to resort to the use of a personal name. This is where the art piece understood as a technique or device of data framing differs significantly from the reidentification demonstration described above, in that it lacks a personal name but it is still capable of configuring a biographical account. Let me here explore the notion of a *nameless biography* further by firstly engaging with some social critiques put forward in relation to the function of biographies in general.

James Clifford has suggested that from an ethnobiographical perspective, 'a person is a sequence of culturally patterned relationships, a forever incomplete complex of occasions to which a name has been affixed, a permeable body composed and

[64] See http://www.filmcomment.com/article/site-specifics/, last accessed 9th August 2015.

decomposed through continual relations of participation and opposition' (1978, 54). As an advocate of ethnobiography and a critic of traditional biographical accounts, Clifford (1978) argued that backgrounds - or 'patterned occasions of experience' (Whitehead, cited in Clifford 1978, 55) - are actually so densely woven into and through persons that one risks doing violence to reality in separating them out, or in trying to define their boundaries. Here, there is a distinction to be made, as Strathern (1992) suggests, between the life that is made to flow or unfold in a ceaseless stream of events, and the punctuated life that acts as a container for a limited number of different occasions set against a static and predefined background.

Clifford (1978) compared the conceptualisation of open-ended and unbounded (ethnobiographical) personhood to what he critically described as the biographical art of arrangement, from which whole and coherent selves emerge. These biographical selves are clearly demarcated from their patterned contexts and described in a diachronic fashion - that is, they are persons and lives portrayed over time. In Clifford's (1978) account, traditional biography is understood as the inescapable fact of personal unity, as a collection of experiences drawn from the broadest pattern of appropriate life occasions. What facts were included and what facts were left out as part of a person's biography, where the context started and where the person ended, was conceived by Clifford (1978) as the outcome of biographical techniques deployed through the craft and method of both biographers and historians of life. Similarly, Pierre Bourdieu (1986) also critiqued such biographical illusion by unravelling some of the presuppositions found in the notion of life histories as 'the sum of the events of an individual existence seen as a history and the narrative of that history' (Bourdieu 1986, 297). In Bourdieu's account, biographies are understood as

significant sequences of events linked to each other on the basis of intelligible relationships, producing the multiple, incoherent 'ordinary experience of life as unity and totality' (1986, 299).

Both conceptualisations[65] suggest particular epistemological and ontological positions with regards to what these authors see as the elusive, contingent and open-ended character of individuals. They both point to how persons are temporarily made up in time, affixed with attributes and characteristics, closed, and circumscribed as wholes through the composition of sequences of events. In summary, both Clifford (1978) and Bourdieu (1986) draw attention to three important elements that contribute, in their view, to shaping the composite arrangements of biographies. Firstly, the longitudinal and chronological order in which life events ought to be organised. Secondly, the juxtaposition and differentiation of persons and backgrounds that needs to occur for biographies to function as representations of personal and unique wholes. And thirdly, the specific connections or relationships that need to be established between parts of a person, and their occasions and events in order for a person to become intelligible as a unit. In light of this, one could argue that in its sequenced arrangement, *I Love Alaska* reproduces the narrative and organisation of the types of biographical accounts that both Bourdieu (1986) and Clifford (1978)

[65] Such conceptualisations also investigate the counter-narratives which compose diachronic lives more densely woven in contexts (or patterned occasions of experience). Both Clifford (1978) and Bourdieu (1986) turned to literature as a format which enables life to emerge as anti-historical, or which reveals collections of events which do not conform to the rhetorical illusion of traditional biography. As an example, Clifford cited William Carlos Williams' Paterson (1963) as an experiment in writing the history of a person and a city on a single plane.

critique: a sequence of life events ordered and linked to each other on the basis of intelligible relationships which bring the 'illusion' of personal unity and coherence.

However, as I have already noted, the biographical account presented in *I Love Alaska* lacks a personal name, and instead features a unique identification number. Techniques and devices like anonymisation bind data's network of connections: in the case of data like search keywords, some relations to the individuals 'out there' - that is, outside of the database frame - are removed, whilst other connections to the world are sustained. The power of anonymisation lies in making the connections between 'flesh and blood' individuals and their queries invisible, without destabilising the data produced by such individuals or threatening their relation to a unit.

Unique identifying numbers allow for such stabilization to occur by maintaining individuals as units in data; this is important as it allows for units to be composed, aggregated and to become commensurable. Unique identifying numbers therefore preserve a qualitative distinction between individuals, while also facilitating their quantification and comparison. So why are personal names anonymised, rather than any other data units? Bourdieu (1986) understood personal names as an institution which integrates and unifies the self. He notes that personal names have the capacity to designate the same object in every possible world - 'in different states of the same social field (diachronic constancy) or in different fields at the same time (synchronic unity beyond the multiplicity of occupied positions)' (1986, 299). Bourdieu argues that proper names are the basis for the unity of a person's successive manifestations, and provide a socially accepted means of integrating such manifestations in a range

of different settings and materials like official records, curriculum vitaes and, in this case, biographies. Personal names have the capacity to amalgamate and organise the relationships between life history events and biological and social characteristics through space and time; they can be thought of as the 'meaningfully situated integrators of social, historical and bodily experiences' (Rabinow, cited in Lury 1998).

However, the proper name in itself does not describe properties and does not convey information about that which it names, as 'since what it designates is only a composite and disparate rhapsody of biological and social properties undergoing constant flux, all descriptions are valid only within the specific limits of a stage or place' (Bourdieu 1986, 300). In other words, a personal name's function is underpinned by a 'cardinal notion of reidentification' (Ricoeur 1992, 32). Personal names not only facilitate the identification of a body as the same in different times and places but also suggest that different diachronic or synchronic occurrences, events, or occasions can be treated as pertaining to the same body under different spatiotemporal schemas. Paul Ricoeur also suggests that 'to identify something is to be able to make apparent to others, amid a range of particular things of the same type, of *which* one we intend to speak' and that it is 'along this path of identifying reference that we encounter the person [with a name] for the first time' (1992, 27; italics in original). The name as a singular denominator makes 'a permanent designation correspond to the unrepeatable and indivisible character of an entity, regardless of its occurrences' (Ricoeur 1992, 29). This is semantically articulated by 'assigning the same phonic chain to the same individual in all of its occurrences' (Ricoeur 1992, 29).

According to Janet Finch (2008), individuality and connectedness are the most important functions of personal names. She points to processes of naming as a way of unificating the "I-" and the "We-Identity" in society: the double construction of the name – namely the forename plus surname formula – 'combines the "I" and the "We" identities of the individual as a way of making visible individuality but also kinship relations to others' (2008, 711). She suggested that whereas 'surnames root the individual in a kin network, the choice of first names provides the opportunity to introduce the dimension of individuality' (2008, 712). Finch further argues that the constitution of personhood through naming is linked to the emergence of the modern state, as the legal requirement to have a fixed name can be traced to the certification of property rights and the need to keep accurate information on individual citizens through birth registers and certificates (2008, 711). Finch (2008) goes on to explore how the doing of names and changes in naming – how children are named; the retention or alteration of a given name; the change of name to a partner's surname - could be thought of as tools for displaying connections, for marking out individuality, and for delineating family relations and genealogies.

Art pieces like *I Love Alaska* not only reframe the framing of search keywords as personal and biographical, but also question whether the force of personal names might be socially less relevant than sometimes presumed- since the reordering of search keywords in *I Love Alaska* flesh out and characterises a person independently of the use of a name to establish relations between individuals (Lury 2005, 95). *I Love Alaska* can be understood as an experimental means of demonstrating that search keywords are part of persons, and that they can become part of biographical

accounts that do not necessarily need the traditional organising force of personal names. *I Love Alaska* is hence an experiment not so only in contexting (Asdal and Moser 2012), but also in the reorganisation of series of search keywords into sequences of words that unleashes a personal but still nameless biographical account. It shows the generative subjectivising and individualising force that emerges despite the deletion of personal names.

The transformations that are enabled by both the database – in its replacement of personal names for unique identifying numbers – and the nameless reconfiguration of search keywords in a biographical fashion can be thought of as processes that make visible 'the dynamic sequencing of framing practices, of connecting and dividing, organized in such a way as to expand modalities of relating' (Lury 2012, 249). These techniques - of anonymising on one hand, and of establishing nameless biographies on the other - reframe relations through which personhood and individuality have historically come into existence through the use of personal names. The extension of relations of ordering and continuity (Lury 2012) permitted by these two framing techniques enables entities like search keywords to emerge as being about persons, and allows biographical accounts to reference personhood without relying on permanent designations, such as personal names.

Conclusion

In this Chapter, I have described different technologies and techniques that configure a particular relation between search keywords and persons. I have shown how search

keywords have become articulated in the operation of search engines as expressions of subjective and unique informational intentionality and needs. I have also suggested that the practice of searching is becoming a mode of equipping persons with particular traits and cognitive dispositions. The Chapter has particularly examined how the relation between search keywords and persons is being defined in social experimental data instances. I have looked at how reidentification demonstrations enable merographic connections to become established between search keywords, persons and their contexts; but also how techniques like reidentificaiton demonstrations can be understood as way of technically redoing data as personal. The Chapter has also paid attention to how search keywords become reconfigured as biographical in the organisation of an artwork. I have claimed that the artwork *I Love Alaska* creates a nameless biography as an expression of personhood, and that this can be deployed to think about the possibly diminishing function of personal names in the context of the emergence of new data types, database practices, and data reorderings.

To conclude the Chapter, I would like to briefly explore whether the traceability of data (Latour 1999, 2010) can be understood as a technical redoing (Goffman 1974), and if so, whether the public traceability of data effected by techniques like reidentificaiton demonstrations can also be understood as a mode of *public forensis,* in the words of Eyal Weizman (2014). I want to argue that, while search queries and many other data types become inscribed as pertaining to a particular moment and setting in space and time,[66] once this happens multiple and partial orderings and

[66] The automation of such modes of inscription in the incorporation of digital time and space data stamps is paradigmatic in this respect.

framings come into play that both extend and bound the relations that data can have with other entities and phenomena in the world, including different spaces and temporalities. Latour (1999) pays particular attention to the scientific infrastructures of referentiality (Lezaun 2006) that enable inscriptions to remain attached to a singular point of origin. The data experiments presented here show that although the trail of connections that link inscriptions with their original settings usually remains inactive - which is one of the ostensible functions of anonoymisation techniques - these connections can become reactivated.

How we conceptualise traceability however changes depending on whether we think of the eventual reactivation of connections within an original setting as the possible outcome of a reversible or irreversible process of inscription. In classical physics, Manuel De Landa explains, reversible processes are usually exemplified by he 'motion of an object in a frictionless medium, such as a ball thrown upwards in a vacuum followed by its downward motion returning to its initial position. A motion picture of this process would look exactly the same if projected in reverse' (2002, 105). He goes on to explain that processes in thermodynamics are not reversible in this sense: systems and their milieus that have changed in certain ways cannot be restored to their initial states; on the contrary, such restoration could sometimes produce contrary or unexpected effects (De Landa 2002, 105).[67] The processes in which the restoration of initial states becomes unattainable are hence defined as

[67] De Landa explains: '(…) most processes in thermodynamics, such as diffusion or heat conduction, are not reversible in this sense. Diffusion, for example, tends to homogenize small differences or fluctuations, that is tends to damp them. But if we reverse the sequence of events we get the opposite effect, a damping process turning into a process of amplification of fluctuations' (2002, 105).

irreversible processes. Analogously, I want to suggest that inscription is not a reversible process, and that therefore that when data is materially re-connected with its it original referent, an altogether different process is set in motion.

It is by thinking inscription processes as irreversible that traceability can be understood as an object of demonstration, or as a mechanism for technically redoing inscriptions in public. As I have already mentioned, Goffman defined and exemplified demonstrations as 'performances of a task-like activity out of its usual functional context in order to allow someone who is not the performer to obtain a close picture of the doing of the activity. [Demonstrations are what happens] when a salesman shows how a vacuum cleaner works to pick up the dirt he has instructively dropped on a housewife's floor' (1974, 67). Rather than thinking traceability as a reversible process of finding the objects and settings from whence data originated, one can think of it as a means through which the staging of the use of technology, or in this case the staging of particular relations between data and persons, becomes publicly accomplished. The demonstration of traceability can thus also be thought as a mode of *public forensis*, as a technique 'for mediating the "testimony" of material objects' (Weizman 2014, 5).

Chapter Four

Ethical Devices and The Economisation of Personal Data

Introduction

In the previous Chapter, I described techniques of data personalisation and how new data types such as search keywords come to constitute new notions of personhood. In this Chapter I concentrate on devices and techniques of data depersonalisation instead. As described in the previous Chapter, initially AOL data was intended to circulate within the confines of a particular academic community. The *read me* file (Appendix, Figure 1.0) which circulated alongside the data stated that the search database was being distributed for 'non-commercial research use only', and that its use for commercial purposes was 'strictly prohibited'. Furthermore, in relation to the possibly sensitive and explicit language contained in the search queries, the same file stated, as a series of caveats, that firstly, the data represented 'real world users'; secondly, it was 'un-edited and randomly sampled'; and finally and more intriguingly, AOL was not 'the author' of the data although it claimed in principle to hold copyright over the data as a collection of units organised as part of a database arrangement.

The data could be made available to researchers working outside AOL because two important ethical principles underpinned the original collection and formatting of the data. Firstly, users had given informed consent either to the data only being made available and circulated for research use or for AOL to share the data with third

parties; and secondly, users' privacy was supposedly not breached as their identities were made anonymous and unidentifiable via the replacement of personal names with unique identifying numbers. Whilst these two ethical principles at first sight seemed to secure the possibility of transferring and circulating the data for academic research as ethical, in this Chapter I claim that ethical principles and the practices through which they become operationalised also participate in the '*economization'* of data (Çalışkan and Callon 2009; Çalışkan and Callon 2010). Whilst the techniques of reidentification demonstrations and data reorganisation frame search keywords as personal, other devices such as informed consent and anonymisation, I argue, secure the creation of economic value out of data by framing it as depersonalised instead.

In the study of the economisation of biological materials, the value created out of the qualities or capacities inherent in life or living entities has been referred to as *biovalue* (Waldby 2000, 2002; Waldby and Mitchell 2006; Rose 2006). Biovalue is described by Catherine Waldby as 'the yield of vitality produced by the biotechnical reformulation of living processes' (Waldby 2002, 310). Nikolas Rose similarly identifies the emergence of economies of vitality, which he describes in terms of the molecularisation of vitality: 'vitality is decomposed into a series of discrete and distinct objects – that can be isolated, delimited, stored, accumulated, mobilised, and exchanged, accorded a discrete value, traded across time, spaces, … in the services of many objectives' (2006, 135). Extending Michel Foucault's concept of biopolitics, the practice of governance that brought 'life and its mechanisms into the realm of explicit calculations' (Foucault 1978, 143), Rose claims that such calculations no longer organise only the state, but instead increasingly shape economic spaces that create value out of life itself. In these processes, the somatic entities that are produced

and stabilised are no longer either individuals or populations, but fragments: cells, molecules, genomes and genes (Helmreich 2008).

Rendering the value of life calculable is, however, a controversial practice, and it is often argued that life resists inclusion in market exchange (Zelizer 1978, 1985; Tober 2001; Scheper-Hughes 2001; Fourcade 2009). This resistance has not however meant that life and its fragments have remained outside market valuation mechanisms; on the contrary, specific devices have served to enable its valuation and circulation (Rose 2006; Karpik 2010). This Chapter explores the possibility that these and other forms of valuing life are not limited to - or exclusively produced in - the unfolding of the bio-sciences within biocapital, but are also implicated in the practices which a range of data scientists and researchers undertake when conducting research with human subjects. Indeed, many economic theories of privacy explore the monetary valuation of personal data. Alessandro Aquisti (2010) for example describes theories that highlight the potentially negative impact on economic efficiency of regulating the flow of personal information; but he also presents theories that offer opposing arguments. The first strand is represented by the work of Richard Posner (1981) and George Stigler (1980), who both considered how regulating privacy as a separate sphere from the marketplace might create inefficiencies on the basis that complete information markets in theory tend to lead to economic efficiency. Regulating privacy conceals potentially relevant information from other economic agents, and it 'potentially transfers the cost of a person's negative traits to other market players' (Aquisti 2010, 5). Aquisti (2010) also points to economic theories which criticise Posner's and Stigler's positions on privacy, and which undermine the assumptions of rational behaviour that underpin such economic models of privacy, arguing that they

fail to capture the complexity of economic decision-making processes (Hirshleifer, cited in Aquisti 2010).

The focus of this Chapter is on how the valuation and marketisation of personal data are accomplished through the workings of ethics devices. The claim I make is that creating value out of life itself has been, and continues to be, a characteristic of the practices of academic and commercial scientists and researchers working with supposedly non-biological materials, or '*vital emissions'*. I think of vital emissions as the movements, actions and dispositions that humans 'release' as part of their daily life, including (but not limited to) searches, utterances, actions, opinions or interactions that are becoming detached, fragmented, recombined and transformed into valuable aggregates with the aid of a range of devices. Across a range of epistemological approaches and disciplines, scientists and researchers have manipulated, qualified and brought together individuals and their vital emissions in order to create an aggregate, like the database, that is qualitatively distinct in value from its individual components.

Indeed I will argue here that a newly enhanced productivity of living entities is emerging, linked to what Mike Savage and Roger Burrows (2007) describe as a new reality of large volumes of digitalised transactional data and data fragments, previously inexistent or considered informational waste, that are being alienated, circulated and reincorporated in increasingly complex economies. Here, value is being created out of what were previously considered human by-products. New technologies appear to be enabling the capture and extraction of vital emissions, such as commercial transactions or movements in space and time, which humans produce

as part of their living process and daily life and which can be fitted into biographical, biometric and transactional classifications, usually in the form of databases. In these practices, the frame of new technologies captures and mobilises vital emissions, and incorporates them into other value regimes (Appadurai 1988): vital emissions are are, in short, being rendered economic (Callon et. al. 2007; Callon 1998; Çalışkan and Callon 2009, 2010).

This Chapter describes how the conversion of vital emissions into data and their subsequent incorporation into databases - that is, the subsumption of the individual into the aggregate - is a highly mediated and complex operation in which ethical devices play a key role. The Chapter explores this mediation of data through a focus on the use of the devices of *informed consent* and *anonymisation,* as a way of showing how ethical practices participate in the creation of value within scientific and commercial research. It demonstrates that one way of creating value within these economies is to convert vital emissions into data in such a way that allows the data to emerge as an entity which is disentangled and distinct from the human agency to which it makes reference. This allows for the circulation of vital emissions in regimes

where they can be made commensurable[68] to one another and rendered of economic value.

In analytically opening up ethics as an economic device, I am not attempting to reveal what lies behind ethical principles or to question their moral implications, but to 'rediscover the strangeness of what has become all too familiar' (Karpik 2010, 10) by questioning the emergence of data as an unmediated expression and function of human agency (Lezaun 2007). In exploring ethics as a device, I consider how informed consent and anonymisation procedures may endow human actors with autonomy and privacy, so as to enable the extraction and manipulation of data. In so doing, I demonstrate that there is no natural, a priori human essence located in research subjects, but instead that humanity lies in the 'entanglements which affect and move those it envelops' (Callon and Rabeharisoa 2004). In other words, what a

[68] It is however worth noting that incommensurability, as a measure which excludes comparisons, is not necessarily incompatible with market practices. On the contrary, as Lucien Karpik (2010) notes, incommensurability is an important asset in what he characterises as an economy of singularities, that is, markets dominated by goods and services which are 'multidimensional, uncertain and incommensurable' (2010, 10). In not accepting the idea of a discontinuity between culture and the market, Karpik argues that under certain conditions incommensurability and commensurability are indeed mutually convertible *within* the market; he says: 'the oscillation between a relatively common stable reality and the multiplicity of the constructions associated with individual and collective points of view is constitutive of markets of singularities. It authorises equivalence without calling incommensurability into doubt' (2010, 10). He further argues that incommensurabilities cannot furnish the basis of a general theory which treats singular products as equivalent entities, that is, which requires them 'to give up precisely that which makes them desirable' (2010, 31). Rather, he argues that that emphasis should be put precisely on analysing the mechanisms by which incommensurability becomes a quality of singular products which are exchanged within the confines of market settings.

‘subject’ is in the term ‘research or data subject’ is the outcome of a specific performance in which ethics play a key role. But ethics do not only bring the ‘research or data subject’ into existence, or make it visible as an entity; ethical practices also support the emergence of something that is extractable out of the research subject, that is, ‘data’ as a separate, objectified thing. Ethics allow for vital emissions (otherwise entrenched in framings of subjectivity and privacy) to circulate and ‘to alternate between detachments and (re) attachments’ (Callon 2007, 343). The Chapter traces the characteristics of a series of generalised informed consent and anonymisation practices, and then goes on to discuss the Data Protection Act of 1998[69] and the Copyright, Designs and Patent Act also of 1998 in the United Kingdom, showing how these legal and ethical principles have been operationalised and how they can be understood as data economisation devices.

[69] The UK Data Protection Act of 1998 came about as a response to the European Data Protection Law of 1995 (Directive 95/46, European Parliament & European Council). At the time of writing and submitting this thesis, the European Data Protection law has been under review. This review has resulted in a Proposed Data Protection Regulation which has still not been operationalised in individual country members, but is expected to be enforced in late 2015 or early 2016. This new regulation will replace the previous Directive and will eventually provide a new mandate for the relation between persons and data for European country members. It is also worth noting that a range of novel and interesting developments have taken place while writing this thesis, like the ruling of 13th of May 2014 by the European Court for the right to be forgotten as applying to search engines.

Informed Consent

Informed consent hashistorically played a central role in ethics codes and is at the core of ethical research practices established by most professional research and scientific codes of conduct. The first and most prominent principle established by the Nuremberg Code (1949) states that:

> The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision. This latter element requires that before the acceptance of an affirmative decision by the experimental subject there should be made known to him the nature, duration, and purpose of the experiment; the method and means by which it is to be conducted; all inconveniences and hazards reasonable to be expected; and the effects upon his health or person which may possibly come from his participation in the experiment. (Nuremberg Code 1949, 181)

The fundamental claim put forward in the Code is that 'informed' or 'freely given' consent by research subjects is a legitimate basis for research, and this has been central to most subsequent regulations governing scientific and commercial research and their data extraction and handling practices (Manson and O'Neill 2007). Both

informed consent and anonymisation principles have been part of the regulation of scientific research practice since the establishment of the Nuremberg Code in 1948, when ten standards were set up for investigators undertaking research with human subjects. Based on these Nuremberg Code principles, The World Medical Association Declaration of Helsinki released a universally applicable set of rules aimed at regulating medical practice in both clinical and non-clinical settings. These two landmark documents have subsequently been developed into multiple international, national and local ethics codes. In the UK, for example, they were drawn on in the establishment of (medical, including biomedical) Local Research Ethics Committees in 1968, and the social sciences in turn arguably imported the principles established in biomedical ethical codes of conduct in order to regulate their research (Beauchamp and Childress 2001). Research with human subjects has thus been organised around a number of self-regulatory standards, which are often adopted by researchers as part of professional codes of conduct. Both informed consent and anonymisation principles have also been incorporated into The Data Protection Act which was adopted in the United Kingdom in 1998 as part of the European Union Data Protection Directive, an initiative mainly aimed at promoting personal data privacy rights and harmonising the data protection laws of its member states. So while research ethics frameworks limited their scope, in principle, to the governance and regulation of scientific research practice, the Data Protection Act became a legally enforceable regulation for any 'data controller', that is, any person or institution that potentially determines the purposes and the manner in which what is defined as personal data is processed. This Act therefore extended the applicability of informed consent and anonymisation procedures to commercial data activities.

Informed consent protects the right of the research subject to be fully informed about his or her participation in any research or data experiment, including possible dangers and risks, as well as granting him or her the right to refuse participation.[70] It sometimes also includes consenting to waive any proprietary claims over the data extracted. The information deemed necessary for consent to be given 'freely' usually refers to the purpose, methods and intended possible uses of the data gathered. 'Informed' consent on the other hand assigns human agents with freedom of choice. In doing so, it assumes that human agents could and should be autonomous from any external sources which might influence their desires, beliefs and decision-making processes. In this regard, the implementation of the practice of informed consent enacts a version of human agency which is sovereign from the influences of other agents, either persons or things.

There are multiple ways of consenting to research or data extractions which can both delimit and expand the uses and the economisation of the data being extracted. Indeed, while informed consent initially emerged as a practice tied to medical research and the human body and is thus particularly concerned with the subject's physical presence and integrity, a range of other types of consenting practices have emerged in response to changing research practices and scientific developments like, for example, genomics (Hallinan and Friedewald 2015). Democratic community consent for example – also defined as 'presumed consent' or 'opt-out consent' – was deployed in Iceland 'to legitimate the collection and use of the Icelandic's population's genetic samples and data' (Hallinan and Friedewald 2015, 6; see also

[70] At the time of writing a number of controversial experiment cases with data without requiring the consent of data subjects have emerged, most prominently Bond et al. (2012).

Arnason and Arnason 2004). This type of consent assumed that unless individuals had actively opted out of the project, they had consented to the future use of their genetic samples and materials. Another recent way to obtain consent from research subjects is 'dynamic consent'. Dynamic consent is designed as

> […] a personalised communication interface to enable greater participant engagement in clinical and research activities. It is a participant centred initiative that places patients and research participants at the centre of decision-making, providing an interactive IT interface to engage with participants. This approach is 'dynamic' because it allows interactions over time; it enables participants to consent to new projects or to alter their consent choices in real time as their circumstances change and to have confidence that these changed choices will take effect (Kaye et al. 2014, 23)

This type of consent enables research subjects to be presented with different projects and up-to-date information so that they can adjust their consent preferences regarding the materials and data collected according to their changing desires and needs. 'Sectoral consent', on the other hand, enables 'a limited extension to the boundaries of specific consent' (Hallinan and Friedewald 2015, 5) in that research subjects consent to their materials, samples and data being used in a general research area, as opposed to consenting to the specificities of any one given research project.

'Open consent' – also called 'portable informed consent', 'blanket consent' or 'general consent' – is a relatively novel type of consent although its history now spans more than a decade (Hallinan and Friedewald 2015). The Personal Genome

Project,[71] for example, an organisation deemed to be one of the main proponents of open consent, states the following with regards to its data extraction and handling practices:

> A surprisingly small amount of data is necessary to uniquely identify a person by name. For example, a zip code, birthday and gender may be enough to infer someone's identity with a high degree of accuracy. One of the most identifiable pieces of information is human DNA sequence: it can be used to predict a variety of medical conditions and traits, like hair and eye color, facial features and even surname.
>
> Because we cannot guarantee privacy and we are committed to sharing data for the advancement of science, we feel the most ethical and practical solution is to collaborate with individuals who are comfortable sharing their data without any promises of privacy, confidentiality or anonymity. We call this "open consent".
>
> If you are considering enrolling in the Personal Genome Project (PGP) it is important that you understand that the data may become publicly identified, by name as yours – even if you exclude your full name or other identifiers like a facial photograph from your public profile. Any hesitation now could cause regret later, after it is too late to remove your data from the public domain.[72]

---

[71] See http://www.personalgenomes.org, last accessed 10th August 2015.

[72] See http://www.personalgenomes.org/organization/non-anonymous, last accessed 10th August 2015.

Open consent presupposes that the research subject actively gives consent only once to a researcher or institution, but that this covers all future research projects and it has no restrictions on whether his or her data, samples or materials are transferred or circulated further.

In the United Kingdom, the Data Protection Act 1998 has taken informed consent as one of its pivotal requirements for enforcing the lawful acquisition, possession and use of what is referred to as either 'personal data', or 'sensitive personal data'. Indeed, informed consent is implicitly required for both personal and sensitive data types in order for the data controller to comply with the Data Protection Act's first principle: that these data types should be processed fairly and lawfully.[73] In the case of sensitive personal data, the Data Protection Act requires explicit informed consent from the research subject, usually established through a written or recorded medium. Informed consent can be obtained in writing through a detailed consent form which usually includes: an informative statement, which can be visually or audio recorded or simply verbal depending on the nature of the research; details of the type of data to be gathered; and the format in which the data will be inscribed and stored.

[73] Further Data Protection Act principles include: 2) Personal data shall be obtained only for one or more specified and lawful purposes; 3) Personal data shall be adequate, relevant and not excessive for the purpose of processing; 4) Personal data shall be accurate and up to date; 5) Personal data shall not kept for longer than necessary; 6) Personal data shall be processed in accordance with the rights of data subject; 7) Personal data shall be kept secure; 8) Personal data shall not be transferred to countries without adequate protection.

Before moving on to examine informed consent as a data economisation device, I want to reflect briefly here on what it entails to think about research ethics principles and operations as devices. Research ethics here are not here considered as textual or discursive representations or renditions of the institutions they were created by. Instead, they are treated as devices which not only enact particular versions of what research subjects are and can do, but also, importantly, affect how data becomes qualified. Understanding research ethics as devices necessitates looking at the workings of specific framings where 'it is very often the discrete, visible or invisible arrangements that work to influence the action' (Karpik 2010, 44). As suggested in Chapter One, Callon et al. (2007) have argued for an understanding of devices as *agencements,* where persons or things emerge from a contingent compound which assigns specific properties to both, as opposed to being already in an a priori 'agenced' state. A device therefore emphasises the distribution of agency between persons and things, and brings materiality to the fore by assigning agency to objects. Devices also position subjects as enacted in and through devices rather than being external to them, so that subjectivity emerges through configurations with multiple devices. In this sense, I understand research ethics frameworks as devices which enable and articulate certain actions, and which facilitate the framing of data as impersonal (or personal). And as a result, as Callon (2007) outlines in relation to the economy, I propose to take research ethics not as a set of moral principles in themselves which regulate an already-existing state of affairs, but instead as instruments which contribute to the construction of data as an economically valuable and viable product by way of its depersonalisation.

While explicitly formulated as a way of recognising the autonomy of individuals,

when understood as a device informed consent can be seen as a means to make explicit the sort of connections or attachments - the sort of dependencies - which produce human activity as autonomous. Specifically, it demonstrates that for consent to be autonomous, it also needs to be framed as informed. Independently of whether it is traditionally face-to-face, community-based, sectorial, dynamic or portable, informed consent practices and regulations demand that information regarding the research process and data processing procedures should always be available in order to allow data subjects to arrive at an informed decision regarding their involvement in research activities. Moreover, the available information should not be deceptive and should not withhold material or methodological facts about the research process. It is therefore assumed that the provision of the right type and amount of information regarding the research process configures research subjects as free and self-determined agents, and endows research subjects with specific capacities which enable them to conduct themselves as particular types of persons.

It can therefore also be argued that the device of informed consent produces the conditions necessary for the emergence of the market ideal of perfect information display and symmetry, which assumes that an appropriate distribution of information between agencies can lead to a legitimate exchange which would otherwise be considered fraudulent or unlawful. Informed consent provides research subjects with enough information to help them make autonomous and rational choices, but these choices are, nevertheless, not possible without the 'preliminary framing not only of the choices, but also of the freedom of the framed actors' (Cochoy 2007b, 115). It can be argued then that it is not autonomy which ensures that an informed decision will be made, but instead the act itself of giving informed consent, mediated by the

provision of information, that produces autonomy and the ability to make free, rational choices by framing subjects in a certain way. It is thus the device that enables giving consent to be rendered an explicitly human form of agency.

Research ethics, regulations and technologies also make privacy visible, insofar as they produce a private sphere that can and must exist before the practices of research and data collection. While privacy is not explicitly defined by the Data Protection Act 1998 as such, its meaning is linked to what is referred to as 'personal' and 'sensitive personal' data. Personal data is defined as data that relates to a living individual who could be identified from the data, or from the data and other information which might be in the possession of, or is likely to come into the possession of, the data controller. Sensitive personal data on the other hand is defined as data which might incriminateor stigmatise a participant or third party, such a s an individual's ethnic origin, political opinion, religious beliefs, trade union membership, physical or mental health, sexual orientation and criminal proceedings or convictions.

Ethics devices create the existence of (potentially extractable) data as deeply entangled in networks of subjectivity, sociality and humanity. The extraction of this data from research subjects is defined in relation to restrictions of access to such substantial attachment. Informed consent thus can be thought of as a device that also paves the way for what would otherwise be considered an illegitimate detachment and transfer of that which is considered personal (private) from a research or data subject to a researcher or data controller. As in market exchange settings, research transactions are framed as consensual (by the device), otherwise they are considered illegitimate acts of theft or fraud. Some examples would be acts in which property is

dishonestly appropriated, without the freely given consent of one party, or where agents are either not given the necessary information or deceived regarding the type of transaction they are committing to. The difference however is that while in market transactions the parties involved are recognised as already holding property rights over what is being exchanged, research or data subjects do not, strictly speaking, hold property rights over their personal data - although they are constituted as subjects possessing legal and moral rights to exclude others from accessing and processing what is considered private.

Under current conditions in the UK, it is in fact indeterminate whether research subjects can establish property rights over their own vital emissions, unless their emissions happen to meet the requirements stipulated in the Copyright, Designs and Patents Act 1988. Under the Act, property rights can be established firstly if vital emissions can be made legible as something materially outside of research subjects; secondly, if vital emissions can be considered authored solely by the research subject; thirdly, if vital emissions can be considered original; and finally, if vital emissions can be considered work. If all conditions are met, then copyright gives human agents 'exclusive legal control over certain acts in relation to their work – not acts of use as such, but only certain acts of replication and repetition' (Barron 2008, 3).

One can think then of informed consent as acting as a 'surrogate property contract' (Waldby and Mitchell 2006), legitimising the detachment and transfer of what is configured as personal into value regimes in which it is protected (Appadurai 1998). Informed consent as device therefore reconciles the epistemic and economic potentiality of vital emissions through their circulation and detachment in the form of

data, at the same time as it preserves the significance of vital emissions as economically valueless and entrenched in privacy framings.

Anonymisation

In the following section I would like to characterise the anonymisation of data understood as a data economisation device. One of the ways in which a division between persons and things is established within Euro-American societies is by means of the assertion of property rights (between persons and with respect to things) 'as the paradigmatic exemplification of ownership – so that when one talks of property ownership one implies that rights are being exercised over [in relation to] some 'thing' or other' (Strathern 2004, 111). Indeed, Strathern further agues that the more entities in question approximate to things rather than persons, the more legitimate ownership appears (2004, 111). In the following section I argue that anonymisation can be understood as erasing previous attachments and facilitating new appropriations and reconfigurations of vital emissions, aiding in the process of making vital emissions more thing-like, and thus more amenable to ownership. This becomes clear when we explore how anonymisation aids in the production of databases (and their ability to be copyrighted): in being anonymised, vital emissions can effectively change from being singular and incommensurable, unique and identifiably about a person, to being data which can be rendered comparable, and may be measured, stored, transmitted, traded and also importantly brought together as a collection.

Let me also here briefly go back to some of the materials that the AOL leak made visible as a way of illustrating how anonymisation becomes operationalised. Figure 1.4 in the Appendix shows the stream of unformatted data produced by AOL's user number 14467039 on the 2nd of June of 2006. If read from left to right it firstly shows an anonymised user id, followed by a search keyword or text, a timestamp, the number in the ranking of retrieved search results clicked through, and finally the hyperlink clicked or accessed via the search interface. AOL's data was originally released in ten text files containing plain text to ensure compatibility and malleability across different database types. The data files released were intended to be fed into various database management systems and be compatible with their specific modes of representation (Castelle 2011). The text files make visible the existence of data as an entity separable from the context in which it initially became inscribed.

In the text files, search queries become standardised, abstract, and formatted as independent of any particular bodily or individual manifestation. When search keywords are disentangled from their users, the latter become amenable to being grouped together but also fragmented in parts in previously unthinkable ways. The disentanglement of inscriptions from their settings is primarily achieved via a series of data transformations and formatting processes, as well as the use of different infrastructures that support both their immutability and traceability (Latour 1999). Figure 1.4 in the Appendix shows that data is not merely a replica or record of a certain user interaction with a search engine. Data has in the first place been formatted to fill tables within a database (this is what the text file enables as a format); it has been cleaned (search queries have been case shifted and any punctuation removed); and data has also been made anonymous (see below). The data

presented in Figure 1.4 is therefore already not 'raw', if there is such a thing as 'raw data' conceived of as a relatively unmediated inscription of some sort (Gitelman 2013).

Confidentiality and anonymity have been deployed in scientific research as two foundational principles in making research ethical, where anonymity is primarily understood as the vehicle by which confidentiality is operationalised. Except in the case of open consent as discussed above, almost all data regulations stipulate that no personal data acquired during the process of research should be disclosed, unless a research or data subject has given specific consent to the disclosure. So in order to enable the circulation of data, anonymisation is typically deployed to remove links that could be used to identify research subjects. This is usually described as a mechanism for the protection of the data subject's privacy. Two main types of anonymisation procedures are generally utilised: reversible anonymisation entails processing extracted data so it does not contain any identifying keys, although researchers retain information that can be used to link the data back to the research subject. Alternatively, data can be unlinked irreversibly, so as the research subject remains unidentifiable not only by other agents but also by the researcher and indefinitely over time.

A number of measures are customarily implemented to ensure that the link between data and identifiable individuals is broken. Datasets, such as AOL's, are anonymised by removing direct identifiers, for example name or address; by aggregating or revising the precision of a variable, for example by replacing date of birth by age groups; or, in the case of geo-referenced data, by replacing point coordinates by

larger, non-disclosing geographical areas which typify geographical position in some way. Anonymising qualitative datasets containing transcriptions or textual data is mainly accomplished by using pseudonyms or replacement identifying characteristics in those data fragments that are considered to reveal personal traits, and by identifying those replacements in a meaningful way, usually in the form of typographical brackets. Voice alteration and image blurring can also make audio and image data anonymous. In those cases where it is a particular configuration of attributes that might enable identification, researchers are required to regroup data in such a way so as to disguise individual identities, or to employ a variety of available measures to impede the detection of identities without inflicting serious damage on the aggregate dataset - for example by disclosing the data at a further aggregate level.

Anonymisation however becomes irrelevant if the actual or potential identification of research participants becomes permissible via informed consent, as the case of open consent clearly shows. Anonymisation therefore becomes unnecessary if informed consent is given to the possible or actual disclosure of personal and private traits. This is seen by researchers as especially desirable in those circumstances in which research participants possess a combination of attributes that make them readily and easily identifiable and where it is difficult to disguise their identity without introducing an unacceptably large measure of distortion into the data.

To give an example, in the United Kingdom anonymisation renders the Data Protection Act 1998 principles irrelevant, as the Act is only enforceable and applicable to data which makes explicit reference to living individuals; in this case, if data is anonymised it is no longer considered personal. As the AOL case discussed in

the last Chapter demonstrates however, the true and genuine anonymisation of data extracted from human agents is difficult to achieve in practice. Weak forms of anonymisation tend not to satisfy the requirements of the Act, whereas stronger forms usually do not satisfy the requirements of research practices which often demand maintaining some traces of the person from the aggregate sets. Anonymisation techniques thus operate at the border of this economy of data: too little anonymisation and vital emissions cannot be converted into data as they are still entangled in the research subject's activities; if on the contrary anonymisation is too effective, then data is stripped of value.

The disentanglement of vital emissions thus requires not only the mediation of informed consent but also the detachment of certain elements, attributes, or marks considered to denote personhood. Only if the products of human living processes can be disconnected from subjects can they become data, and start to be stabilised as thing-like and therefore, own-able. Like body parts and tissues, vital emissions, if understood as an extensions of the self, need to take on a separate existence from that of the original person to which they are linked (Waldby and Mitchell 2006), a separation that is enabled by anonymisation. Furthermore, as well as making vital emissions more thing-like, anonymisation can also be thought of as a technique that suppresses the 'author-function' (Foucault 1970), a function concerned with sustaining conceptions of the individual in terms of self-possession and subjectivity. It was through 'the author-function that cultural value became a thing, a product and a possession caught in a circuit of property values' (Lury 1993, 23). However, within the economy of personal data, it is not the association, but instead the suppression of signs that index a person's existence, that produces a similar effect. That is, such

suppression enables the detachment and circulation of vital emissions in regimes where they can be rendered appropriable.

This contrast between the 'author-function' (Foucault, 1970; Lury, 1993) and anonymisation is worth exploring further. Foucault (1970) describes a particular way of being a person in relation to an idea, work or sets of text as the 'author function'. He presents a socio-historical analysis of the emergence of the author as an individual, and the function and configurations that enabled a text to point to an author as outside and preceding it. The constitution of property rights over intellectual work has been underpinned in Western societies by particular notionsck on my of authorship. An author is, in the modern sense, 'the creator of unique literary or artistic works, the originality of which warrants their protection under laws of intellectual property' (Jaszi and Woodmansee 2003, 195). Recently databases have come to be protected under European copyright Law[74] on the basis that they might be treated as a distinct class of creative and original work because of the particular selection and arrangement of their contents. For copyright to be applicable the arrangement of the contents of the database must be considered original, as in the case for example of literary or musical compilations.

In the economy of data, what effectively enables the conversion of the inalienable into a thing and therefore potentially into an object of property claims *is not the presence of a personal name – as it is typically the case with the author function - but its deletion*. In this movement, value is redefined 'not in terms of the relations of personhood but of relations external to the person' (Adkins 2005, 119). Vital

[74] See Directive 96/9/EC.

emissions become valuable not because they are fashioned by a creative and unique self, or rather, an author, but because they are *potentially* detachable, malleable and convertible into a thing - data - that needs to be (re)animated in order to acquire a value that ceases to be personal.

Anonymisation here performs an important function for the emerging economy of data: it enables the transformation of vital emissions, located in the private domain and qualified as singular (as possessing unique properties) and incommensurable (as not equivalent to any other), to data which in fact *can* be aggregated, compared and rendered commensurable with other data units - even if those units were originally collected from unique research subjects. Anonymisation makes the manipulation of vital emissions possible, and allows for their reordering, classification and grouping under databases, potentially protected in the United Kingdom under copyright law. It is interesting to note here that 'single' facts do not have copyright protection, but that collections of facts may enjoy copyright protection as literary works, databases and compilations which, by virtue of the selection or arrangement of the contents, are considered to be the data controller's own intellectual creation. This effectively excludes the individual as a recognised source of value.

Anonymisation thus aids in the emergence of the database, and in the construction of what Evelyn Ruppert defines as 'categories or classes of equivalence through which individuals pass from their singularity to a generality' (Ruppert 2009, 7). As Ruppert notes, generalising individuals into the population partly involves 'classifying and identifying her difference and resemblance to numerous categories (male, female, married, single, etc.)' (2009, 7). However, the passage from singularity to generality

entails and requires the omission of certain categories, such as personal names, which might allow the original individual components to be traceable. That is, categories which render the elements of the aggregate too individual need to be omitted, as they might make the generalisation of the individual into the aggregate problematic.

The argument being made here, then, is that data is not granted value by the sole fact of having been detached. If data remains detached inert and isolated, that is, if it is preserved as a singular data unit, intellectual property rights cannot generally be awarded to either the data subject or the data controller. In the creation of value, the act of processing and bringing data units together under a single collection becomes key, as it is in the practices of 'animating' the data extracted that potential economic value resides. So, for example, the Data Protection Act makes clear that processing data not only means obtaining, recording or holding data but also carrying out an operation or set of operations on the data. These include the organisation, adaptation or alteration of the data; retrieval, consultation or use of the data; disclosure of the information or data by transmission, dissemination or otherwise making available; and alignment, combination, blocking, erasure or destruction of the information or data.

The passage from individual to aggregate enabled by anonymisation thus creates value by provoking not only a disentanglement of vital emissions, but also by invoking a categorical transformation of vital emissions, a transformation necessary for the production of the aggregate. The depersonalisation of vital emissions renders them comparable and commensurate to each other; through this process, they can be aggregated and therefore become valuable. Singular, non-aggregated vital emissions

are economically valueless; they must, rather, be produced as an abundant resource, highly reproducible and easily replicable. The requisite for vital emissions to become economically valuable under research and data extraction and processing settings is, paradoxically, that they need to be encountered and qualified as singular/personal in their mode of production and then lose (some of) their singular/personal - and human - qualities in the process of becoming economically valuable.

Conclusion

This Chapter has explored the capabilities of the ethical devices of informed consent and anonymisation in the articulation of a particular kind of data economy. It began by showing how informed consent enacts a transaction between data subject and data controller in terms of legitimacy and enables that transaction to become the basis of a surrogate property contract (Waldby and Mitchell 2006). It went on to describe how anonymisation facilitates the conversion of vital emissions into data by suppressing the author function, enabling the rendering of singular data as units commensurable with one another, and therefore able to be aggregated in the form of databases - a procedure that also supports a specific set of intellectual property rights. I have also pointed at the need to reflect upon new ethic devices in the light of contemporary research practices, the development of new technologies and the extraction of different and more inscribable vital emissions.

In the Chapter I have also shown how ethic devices like informed consent and anonymisation contribute to framing data as impersonal; how they establish a suitable

frame in which a particular type of exchange can take place; and how they strip data of relations which might indicate personhood through the deletion of personal names. These devices frame and economise data. By stressing the framing of data as impersonal and therefore economic, the Chapter has aimed at identifying - in the tradition of STS, which proposes a performative theory of the economy - the frames and devices which constitute 'the behaviors, organization, institutions and, more generally, the objects in a particular society which are tentatively and often controversially qualified, by scholars and/or lay people, as "economic"' (Çalışkan and Callon 2009, 370). This has entailed thinking the economisation of data as an achievement rather than as a given or pre-existent reality that can simply be revealed and acted upon (Çalışkan and Callon 2009, 2010). This achievement is accomplished by the depersonalisation of data via anonymisation, a process that secures a particular attachment of data to bundles of property rights and which makes it possible to make particular ownership claims on data. On the other hand, both informed consent and anonymisation secure a stable disentanglement of data from framings of privacy and personalisation and make it 'less vulnerable to the constant pull of re-entanglement' (Çalışkan and Callon 2009, 7). Through such devices data undergoes a process of standardisation that, as the Chapter has shown, is also crucial for its economisation.

As a way of concluding the Chapter, I want to reflect on the future of ethics devices for the social sciences in relation to the framing and economy of data. I would like to indicate two important directions for these reflections. On one hand I want to suggest that the fragmentation of data and the intensification of what is extractable as data from human agents is in line with what might be called the *molecularization of vitality* (Rose 2006). This should, I argue, lead us to rethink ethics as a device that

creates a disaggregated 'persons' that become distributed within new datascapes. On the other hand, I also suggest that the principle of *traceability* (Latour 1999, 2010) might usefully be developed to redefine the devices that sustain value production, and in this way potentially alter how value is produced within data economies.

The economy of digital data traces has recently been portrayed as a fertile empirical terrain over which social science, and in particular sociology, should claim some jurisdiction (Savage and Burrows 2007; Latour 2010). For Savage and Burrows, for example, digital data presents both a challenge and an opportunity for a discipline like sociology that is seeking to redefine itself through a concern with research methods, 'not as particular techniques but as themselves an intrinsic feature of contemporary capitalist organization' (2007, 890). Latour's concern, by contrast, is with the forms of traceability that digital data offers to the researcher (2010). He aims to develop techniques of traceability that debunk the aggregate, the law or the structure as a totality, and to explore the possibility of social research methodologies and devices that produce provisional visualisations of the 'whole', 'which can be modified and reversed at will, by moving back to the individual components, and then looking back at other tools to regroup the same elements into alternative assemblages' (Latour 2010, 116).

Latour (2010) argues that what sociologists have tended to do is to observe and identify the individual and the aggregate as two separate and incompatible entities: the more proximate one is to an aggregate, the less individual variations one can trace. New datascapes, devices and techniques, he suggests, allow for a redefinition of the relationship between the 'ingredient and the structure', in which the aggregate

loses ‘the privilege it maintained for one century’ (2010, 14). In so doing, they enable the development of empricial tehcniques - traceability - that follow individual components in such a way so as to produce highly unstable aggregates, which may be continually dis- and re-aggregated, enabling the researcher to map multiple trajectories in relation to a variety of assemblages.

In both cases, the new reality of digital data traces expands that which can be measured and accounted for by sciences whose object of study are the relations between individuals and aggregates (parts and wholes). For Latour, thanks to new technical developments, the principle of traceability is already being put in motion for ‘opinions, rumors, political disputes, individual acts of buying and bidding, social affiliations, movements in space, telephone calls, and so on’ (2010, 116), while Savage sees the potential for ‘descriptive assemblage’ to be ‘dramatically enhanced by the infrastructure of information technology and more particularly the digitalization of social relations’ (2009, 155). This expansion of what is measurable in social scientific and commercial research produces, on one hand, a reorganisation of the elements and boundaries of sociological units of analysis; and, on the other, due to the scale and diversity of increased traceability, a redefinition of how the trajectory from the individual to aggregate can and should be understood (Latour, 2010).

One can start by thinking of social data units as being redefined as new technical advancements that facilitate an unprecedented and rather uncomplicated multiplication of data. This expansion and intensification of what it is possible to extract from human agency, and the relationships that can now be established

between different data points, destabilises both the individual and the aggregate as the paradigmatic units of analysis of social investigation. It is possible instead to trace the constitution and trajectories of 'monads', that is, entities that are 'a representation, a reflection, or an interiorisation of a whole set of other elements borrowed from the world around it' (Latour 2010, 116). In his reading of Gabriel Tarde, Latour rejects traditional sociological understandings of the 'individual' as a whole, composite site of social investigation, and instead proposes taking monads and their relationality as units of analysis. As Latour says, 'if there is nothing especially structural in the 'whole' it is because of a vast crowd of elements already present in every single entity' (Latour 2010, 10): new datascapes make visible such relationality, and constitute such relations as sources of value.

However, what I want to suggest is that if such new data realities produce a redefinition of what could be considered sociological units of analysis (or what the object of sociological enquiry should be), it should present a challenge *not only* to traditional methodologies and analytical frameworks (Savage and Burrows 2007; Law and Urry 2004) but also to the current governing ethics frameworks, either professional or legal, which regulate and bound the practice of the social scientific but also commercial research enterprise. The emergence of such new units of analysis destabilises fundamental ontological considerations which are naturalised in most ethics regulations: is data 'personal' when actions, utterances, traits or bodily descriptions are not physically joined or found in or through the individual? Is the individual the source of data when he or she is multiplied in a conglomerate of extraneous self-duplications? Is the individual or the society to be found in the action,

the purchase, the gesture, the location, the time frame, the opinion, the conversation, the discourse?

While current ethical regulations may be applicable to a conception of the person as an individual whole that is indivisible, singular and private, it is debatable whether they are still relevant in a context where the distinction between the public and the private is being blurred and where the signs that denote personhood are changing. In this context – that of what Strathern once described as one of the *dividual* (1988) - ethics understood as the 'application of moral principles to concrete social facts' (Fraser 2008, 14) appears to be emptying out, its function becoming predominantly economic. It becomes important then to *rethink ethics for societies* of the *dividual* or post-human, for research subjects who are constituted by the action, the purchase, the gesture, the location, the time frame; to rethink ethics for data which are no longer envisaged or enacted as static, but mobile, constituted in traces, fragments potentially valuable in their futurity, transformed and valuable in the processes of dis- and re-aggregation. Following Mariam Motamendi Fraser, this is to propose rethinking the ethics of the virtual, that is ethics not solely for 'entities that are physically or conceptually present somewhere (just not here), but to virtual multiplicities or singularities that have no corporeal presence at all' (2008, 19).

At the same time, if ethical principles can indeed be thought of as economic devices, as the Chapter suggests, it also becomes necessary to rethink their capabilities in a context where there is a more generalised mode of value production emerging within social and commercial data research. If the aggregate or structure is 'what is imagined to fill the gaps when there is a deficit of information as to the ways any entity inherits

from its predecessors and successors' (Latour 2010, 11), traceability puts the focus on circulation, rendering the tracing of the *trajectory* from the individual, to the aggregate, and back that which becomes central. What might create value here are not any of the provisional steps, that is, from the individual or the aggregate taken separately and independently from each other, as informed consent and anonymisation principles currently contribute to produce, but the trajectories that constitute the individuals and aggregates as processes and not as things in themselves. In this sense, if traceability, as Latour suggests (2010), should become a new method or technique deployed as part of scientific and commercial research practices, there needs to be a consideration of what types of ethics such practices require - and what types of economic value they would promote. The creation and design of the range of devices which intervene in the production of data economies and the type of value forms they enact are too crucial to be overlooked.

Case Study I - The Making and Unmaking of Personal Data

Conclusion

This case study has engaged with both newly-emerging and long-established techniques and devices that frame the relation that different data types can have with persons. In other words, I have claimed that techniques and devices operate as part of a diverse range of practices that frame and configure what data is about. As I have shown in Chapter Four, the relations that these techniques and devices afford in turn contribute to the emergence of a particular data economy. However, as I have described in Chapter Three, they also partake in new modalities of doing and signalling personhood. The techniques and devices I have presented establish a type of relationality that can be conceived of as a form of 'hyperactive kinship' (Seaver 2015, 32), that is, a type of relationality that enables data to become about multiple entities and phenomena in the world whilst also shaping and drawing boundaries.

Throughout this case study, I have elaborated on what makes a strip of data, like a search keyword, personal or impersonal. I have described how web-based search engines personalise search and how they articulate users' search presumptions based on the configuration of data histories and contexts, alongside a range of other analytic operations. The ways in which search engines entangle keywords with persons, and how they shape personal cognitive dispositions in the process, open up and simultaneously delimit a particular mode of experiencing personhood through search. Other techniques of personalisation, such as biographical accounts, also open up

modes of experiencing personhood through the re-organisation of life events. Following Hacking (1986), it could be argued that techniques of personalisation like the ones described in this case study are dynamic, in that they not only act upon that which they intend to personalise but in so doing also ensure it emerges as real.

In this case study I have also paid particular attention to the techniques that frame the relation between persons and data as active, and sometimes deliberate, interventions enabling contextualisation. The framing of search keywords, through their reinsertion into a private contextualisaiton, configures them as personal data, and as such as an expression of personal interiority and intimacy. The reidentification demonstration of Thelma Arnold established a merographic connection (Strathern 1992) between data and a context, a movement that allowed the context to illuminate search keywords as a part and expression of a person's subjectivity. On the other hand, technical redoings (Goffman 1974) can also be thought of as techniques that, by mobilising, staging and enacting a particular context, bring about a different organisation of the meaning and experience of techno-scientific objects (in this case of search keywords). I have also argued that ethical devices contribute to making visible and framing privacy contexts, as ethics regulations and technologies configure a private sphere deemed to exist before the practices of data collection and extraction. Data is framed as existing already-entangled in subjective and private frames. Its extraction or disentanglement from such frames is predefined in relation to the restrictions of access to such data attachments. That is to say, ethical devices on the one hand sustain the framing of already-extracted data as separate, and on the other frame its extraction at a particular point of origin.

I have also engaged with the function and deletion of personal names as a device used in the framing of data. I have firstly looked at the configuration of a nameless biography through a data art work. I have suggested that what this artwork, based on the reorganisation of search keywords, contributes to indicate is the demise of the function of personal names as organisers of relations between individuals and biographical data. I have also looked at how the device of anonymisation enables vital emissions to become detached from privacy framings and be reconfigured by other frames, like the database - rendering the data economic through copyright law. In the economy of personal data, it is not the association that personal names have traditionally established with things - as in the relation enabled by the author-function (Foucault 1970) - which shapes data as an economic object, but the suppression of personal names instead.

This raises a number of different questions with regards to the political economy of data, and the displacement that data subjects and research participants face as authors, when anonymised data becomes compiled and organised as part of databases. Bill Maurer (2015) has recently argued in this respect that kinship theory in anthropology might provide alternative resources to speculate on the formation and shaping of data economies. He points out that what is interesting about the economy of data are the new relations – of descent authorship and alliance – that such an economy configures and redefines. He claims that in the emerging data economy, questions of data ownership and property need to be thought side-by-side with questions regarding the kinds of relations data can afford, and the types of entities and settings data becomes entangled with. The sorts of property claims that data subjects might be able to make

concerning personal data depend upon what kind of connections data is configured to have, that is, what types of relations the framing of data might enable and constrain.

Traceability has been another key conceptual theme developed as part of this case study, in order to account for the formation of personal or impersonal data. Traceability in this case is the operation of tracing data to a point of origin, a process that was clearly undertaken in the reidentification of Thelma Arnold, but which is also being deployed in the practices of social scientists working with data traces who, 'thanks to digital traceability no longer need to choose between precision and scope in their observations' (Venturini and Latour 2009). Tommaso Venturini and Bruno Latour (2011) claim that for social scientists, 'it is now possible to follow a multitude of interactions and, simultaneously, to distinguish the specific contribution that each one makes to the construction of social phenomena' (2011, 5). I have claimed that inscription should be understood as an irreversible process. This enables us to think of traceability - a process of reverting data back to a point of origin – as a demonstration and an event through which inscriptions become technically redone. If undertaken in public and for a public, traceability can be understood as a mode of public forensis (Weizman 2014); that is, as a technique deployed to mediate and shape the history and biography of material objects like data units. I have also suggested that the principle and analytical application of traceability has the potential to change how value might be produced within new datascapes. This is because traceability debunks the traditional relation established between the individual and the aggregate - a relation that devices like informed consent and anonymisation have historically contributed to configure.

I have also suggested that traceability requires thinking about the current relevance of ethical regulations that assume 'individuals' rather than 'dividuals' (Strathern 1998). The conception of the individual is that of an indivisible self or person. Biographies and personal names are functions of this conception of a person understood as singular and essential core, a singularity which as a whole contributes to 'define a self in its particularity' (Smith 2012, 53). By contrast, 'dividuals' are considered to be divisible, comprising a complex of separable – both temporally and spatially – interrelated but also independent parts, dimensions or aspects. This is similar to what Clifford (1978) argued that ethnobiography, understood as a method, should attempt to capture. Whilst the individual is conceived of closed and given, the dividual is instead fractal - that is, composed and traversed momentarily in and through a multiplicity of interactions and contexts. Strathern argues that persons as dividuals are 'frequently constructed as the plural and composite site of the relationships that produce them. The singular person can be imagined as a social microcosm' (1998, 13). The ethical regulations described in this case study are based on a conception of the individual as singular and indivisible, whereas new datascapes and the development of new analytical techniques such as traceability require the design of ethical devices for persons understood as *dividuals* instead.

I have also shown that it is not only persons that should be thought of as fractal, but also the distinction between private and public. It could be argued that research ethics regulations assume that the private and public domains are a priori separate spheres of activity, attributed with given qualities. However, vital emissions do not happen in already predefined spheres of public or private activity. In fact the distinction between public and private is continuously organised through what Susan Gal (2002)

defines as fractal distinctions and re-calibrations. In Gal's account, 'public' and 'private' are not particular places, domains, spheres of activity, or even types of interaction, and 'even less are they distinctive institutions or practices' (Gal 2002, 80). Rather, public and private distinctions should be looked at from a material-semiotic perspective that examines how the combination of particular discourses and communicative acts, alongside material configurations and devices, articulate such distinction in the first place. Gal (2002) however does not conceive of the production of these distinctions as a single dichotomy established in the operation of a single boundary which splits the world into two spheres or domains: that of the private and that of the public. She suggests instead that the distinction between public and private unfolds as a process of recursive fractal differentiation. The term fractal, Gal (2002) tells us, is used in geometry to describe how a single pattern recurs inside itself as part of multiple nestings. Such a conceptualisation of the enactment of public and private spaces, occasions, and entities, implies that what is at one point fractally calibrated as private can, due to the workings of other calibrations, become public instead. We should therefore also think about what type of ethical devices might be able to operate at the active borders of the fractal distinctions that actively produce and differentiate the private and the public, and the personal and impersonal, as part of everyday social life.

Case Study II - Price Formations and Statistical Compositions

Introduction

The previous case examined a range of devices and techniques that configure search keywords and other data types as personal. In this case study, I look at the framing devices that stabilise a different and perhaps more traditional data type instead: that of *price*. I do so by looking at how methodological and technical devices and techniques deployed in the measurement of different offline and online inflation procedures contribute to forming and bounding prices in particular ways. Whereas the previous case study focused on the leak of search keywords, this case study is based on the request made by government officials in 2006 to publicise granular inflation data produced at the National Institute of Statistics and Census (INDEC) in Argentina, a request which triggered a controversy over the validity of the Consumer Price Index (CPI) and other indicators in the country. The debate over the public release of national statistical databases and data revolved specifically around the degrees of specification and identification that such a data disclosure and disaggregation exercise would afford. Although government officials requested a full disclosure of the products and settings used to calculate inflation in the country, the statisticians involved refused to disclose these details, as they believed that the public disaggregation of data could interfere with the objective measurement of CPI. The statisticians argued that the data should not be released because it would then be at the disposal of government officials, who might subsequently negotiate price agreements with the producers of the specific goods and services that are used to measure inflation. The concern was that any such agreements could destabilise the

objectivity of CPI, which was designed to represent the pure price variation of different products and services 'in the wild' (Callon and Rabeharisoa 2003).

The debate that started the controversy over the validity of CPI and which later spread to other indicators thus revolved around the extent to which itemised data used to calculate inflation should be made public. On the 3rd of July 2007, a *Statistical Secret Act Committee* was convened by the then Director of INDEC to decide whether the institute should grant the request by the government to disclose the databases. The Committee arrived at the conclusion, in accordance with the Act (Law number 17.622 article 10), that the data could only be disclosed if a) it did not enable the identification of price informants, be these physical location(s) or person(s); and b) the disclosure of data did not allow for the identification of a product's brand or reveal the actual (real) object or substance of measure. The Committee claimed that

> The data units that conform a sample – businesses, individuals or households – must remain under the protection of the Statistical Secret Act. Under this legal protection, members of the national statistical system are not allowed to disclose the identities of the data units that conform the sample, even when third parties who do not abide to the Act should request these. When the data that compose the CPI is requested in detail, this can be offered as long as its disaggregation does not enable the actual identification in the field of branded products or informants. (Pedido de Indagatoria 2007, 37; my translation).

The Committee also concluded that the data should not be disclosed if such revelation would enable the transformation or modification of the relations that products and

prices hold naturally -that is, without the influence of governmental or other external interventions. Indeed, the international and local manuals[75] that document the methodological procedures required to measure price variation specify that the annotation of prices must be independent of governmental price control mechanisms, temporary offers, sales and price reductions, amongst other factors. This is so that the result of such measures is a variation of price understood to be natural and pure, expressing only the result of the balancing and coordinative mechanisms of supply and demand. The Committee formed in Argentina concluded however that the data requested *could* be made public if it did not provide information about the particular branded products used to calculate CPI. That is, information relating to a variety of generic products, for example *Brown Bread*, could be disclosed but only if it did not specify - usually via a brand name or detailed description - the particular *Brown Bread* being taken into account to calculate national inflation.

I take the decision reached by the Statistical Secret Act Committee as an apt introduction to this case study because it shows one legal device in operation - the Statistical Secret Act - that served to secure and sustain a particular relation between prices and consumer goods. This case study examines the framing devices and techniques that configure what prices are about, how they vary, and the relations that they establish with a range of different global phenomena and entities related to price variation and inflation measurement. In the example here described the Statistical

[75] See for example Instituto Nacional de Estadística y Censos (INDEC) (2001), Indice de Precios al Consumidor Gran Buenos Aires base 1999=100, Metodología Número 13; Office of National Statistics (ONS) (2012), Consumer Price Index, Technical Manual; International Labour Office (2004), Consumer Price Index Manual – Theory and Practice.

Secret Act is a legal device that prevents the disaggregation of data from acting as a means of identifying retail products and the relations they have with prices. The Act prevents the data from revealing the framing operations of the retail spaces that configure the relations between products and prices. I argue that how prices relate to entities and phenomena in the world, and how their fluctuation is observed and interpreted depend on the operations of framings and the devices and techniques which make these frames possible, visible, intersected and superimposed. The case study explores by taking up the empirical occasion presented by the controversy in Argentina over the validity of the CPI, the different devices and techniques which are being methodologically, legally and commercially deployed to define the *aboutness* of price, to stabilise its formation and to secure its different particular forms of variation.

In following the events, materials and new entities that were brought forth as a consequence of the attempted disaggregation of statistical data in Argentina and the controversy generated over the validity of CPI, I also intend to render visible different framing techniques and devices in order to understand how both offline, online, analogue and digital statistical compositions and decompositions are implicated in the formation of price. The first technique I analyse is the *product identification procedures* undertaken in order to measure inflation. The identification procedures that govern which products and their prices are used to measure inflation are a series of instructions which price collectors follow in order to stabilise the variation of price as pure. This is achieved by holding constant the qualities of the products to which prices are made to refer to as part of retail spaces. I then turn again to examine the Statistical Secret Act as a device that makes the superimposition of price framings

invisible. Finally, I analyse two devices that enable the economisation and financialisation of retail price by facilitating its extended contextualisation. The first device is *scraping*, a digital method for the automatic capture of web-page data, which analysies their code and extracts from the code the relevant data required to stabilise the variation of prices. The second is *imaging*, a device used to contextualise the variation of price and to associate price variations with phenomena like stock availability and inflationary emergencies.

This case study is comprised of two Chapters. The first Chapter – entitled *The Methodological Formation and the Secret Containment of Price* – starts by describing how the material and semiotic formation of price has been conceptualised in economic sociology and the social study of finance. I particularly focus on how the framings of prices in retail spaces have been analysed as mechanisms for eliciting consumers' cognitive dispositions towards price, and how these framings are not singular but multiple, operating simultaneously with other frames. The second section of the Chapter focuses on the technique used to identify the goods taken into account for measuring inflation. I demonstrate that this technique is a framing of price that is patterned on still different to the framing that occurs in retail settings. In the third section, I discuss how the superimposition of frames - or *laminations* (Goffman 1974) - can become visible, and point to the devices which enable or prevent such visibility. I do so by analysing the role the Statistical Secret Act played in this particular request for disaggregated inflation data in Argentina, and claim that, in this case, making the superimposition of retail price and inflation frames visible would have entailed the production of a further frame lamination.

In the second Chapter – entitled *Digital Methods and the Financialisation of Retail Price* – the case study moves on to investigate some of the devices and techniques used to frame digital prices. The Chapter starts by describing the differences in function of aggregated statistics and the price system as described by Friederich Hayek (1945). It argues that the function of economic statistics is changing due to the introduction of data recording and communication technologies which enable more economic registers to become available, and that facilitate the establishment of new associations between aggregated and dispersed data. I claim that these changes might also imply a shift in how the economy is observed, how its fluctuations are interpreted, and therefore how economic action is coordinated. I explore these changes through the analysis of two different devices for framing data that have emerged as part of the development of digital methods to measure economic change - some of which have come into existence in fact as a consequence of the controversy generated over the validity of the CPI in Argentina. The first device is that of scraping, used to extract data out of web pages. I claim that the commercial scraping of prices for the measurement of inflation not only presupposes an extended contextualisation of price but also reconfigures the social function of public statistics like CPI. The second device, that of digitally imaging the contexts of price variation, frames the detection and exploitation of price emergencies as a local knowledge problem. The Chapter concludes by arguing that the digital framing of price contributes to the financialisation of retail price, that is, it shapes the observation and interpretation of retail price variation for financial purposes.

Chapter Five

The Methodological Formation and the Secret Containment of Price

Introduction

This Chapter explores some of the devices and techniques that are deployed to select the prices that are used to measure inflation, and to make prices account for variation in time. It also looks at the devices which prevent different price framings from becoming visible when superimposed. Here, I examine the *lamination* of frames (Goffman 1974), that is, the layering of different frames that at any given time operate to define what (retail) prices are about. I then explore what is politically at stake when different price framings become visibly entangled. I pay particular attention in this Chapter to how these different laminations become observable or unobservable for different agents, and the implications of this for the formation of price *as about* a range of economic phenomena. How prices are experienced, sensed, observed and paid attention to depend upon the material and semiotic operation of frames; these frames define what prices are about, and delimit or expand how price can become related to different objects, settings and temporalities in specific ways. One important technique for framing prices (and price indexes) is their presentation in a composite arrangement and in a text-based format. This is a form of disclosure that does not reveal the forces and relations that enable prices (and price indexes) to be fixed at a given point in time. As such, Jane Guyer (2010) claims that this constitutes a mode of framing prices' moral and political constitution.

Sociology and anthropology have developed different approaches for conceptualising the function of prices in society and for undertanding how 'the social' - understood in a variety of different ways - affects the composition and formation of prices. Jens Beckert (2011) has recently reviewed a range of sociological and anthropological approaches that have conceptualised *price formation* from a social, as opposed to an economic, point of view. Beckert (2011) bases his analysis on the price theory developed in the work of Emile Durkheim (1947). Beckert claims that Durkheim conceptualised prices in two different ways: firstly, in line with economic price theory,[76] he thought of prices as something external to market agents who confront prices as if outside of these. Secondly, and as opposed to economic price theory, Durkheim considered prices as 'reflecting public opinion on the value of a good; prices correspond to the normative principles for a just allocation of goods' (Beckert 2011, 2). Beckert (2011) points out that Durkheim here diverged from economic price theory by understanding prices not as the aggregation or condensation of individual preferences, but as the aggregation or condensation of social norms and value judgments. Beckert goes on to propose that whilst it is questionable whether prices can be understood as the reflection of aggregated opinions of what society deems just, the task of sociology is indeed to account for the social and political forces that intervene in the formation, stabilisation and destabilisation of prices.

This Chapter proposes to study such forces from a framing perspective. Specifically, it will investigate the function of different framing devices which work to define and

[76] Beckert (2011) here refers to economic theory which posits stipulates that prices (and the price mechanism) emerge as the outcome of supply and demand forces and the configuration of economic agents as rational and autonomous.

stabilise the aboutness of retail prices. My understanding of *price framings* has been particularly informed by a branch of economic sociology dedicated to the study of the semiotic-material arrangements that enable price formation and stabilisation (see in particular Muniesa (2007)), and the social configurations that make possible the realisation of price in practice (Caliskan 2010). Within economic sociology, and particularly in the social studies of finance, others have also studied the technologies, contexts and networks that enable a distributed, social apprehension of prices as part of financial markets (Beunza and Stark 2012; Prato and Stark 2013; Beunza et al. 2006). The object of these studies has been to describe how market agents pay attention to, are aware of and attribute meaning to the movement of prices in financial settings with the support of devices and prostheses that make forms of price cognition emerge as socially interdependent.

In these accounts, cognition is not understood as individual, a priori or ingrained in the mind of market agents, but instead is thought of as emergent and immanent to the relations that different devices - like screens, mathematical formulas, models and algorithms - articulate. The cognition of price - the capacity of market agents to read, experience, and interpret prices and their fluctuations in particular ways - has either explicitly or implicitly been formulated in these accounts as prosthetic, that is, as enabled and mediated by the deployment and configuration of a range of devices and techniques (Lury 1998). But this mediated cognition has also been understood in these studies in a social manner, in that devices and prostheses are understood to enable a mode of articulating the observation, attention, and meaning attributed to prices, thus bringing together collectives constituted by a multiplicity of market agents.

In this Chapter I follow a similar set of concerns with regards to the material-semiotic stabilisation and formation of price, which, following Goffman (1974) and Callon's incorporation of an STS focus on material devices, I refer to as the *framed organisation of price aboutness*. Here, however, I depart from the empirical focus on markets (of different kinds) that economic sociology and the social studies of finance have so far relied on. Instead, I explore the framing of prices as part of the methodological procedures and legal protocols used to measure price variation and publicise it as inflation, normally in the form of Consumer Price Indexes (CPI). The first section of the Chapter offers a conceptual review of what I call *the aboutness of price.* I go on to trace some of the work undertaken in the social studies of finance and economic sociology that has looked into the framing of prices, taking such framings as both material and semiotic. In this first section of the Chapter, I also offer a description of how retail prices in particular can be understood as framed by different devices that enable a form of cognition understood as *qualculation* (Cochoy 2008; Callon and Law 2005; Callon and Muniesa 2005; Thrift 2004). In the subsequent section I then move on to describe some of the price framings that become articulated in retail settings where CPI measurements take place, focusing on the identification and specification techniques for stabilising price variation as pure. The final section of the Chapter looks at the legal device that prevented the public leak of the prices and products which had been used to measure inflation, as part of the Argentinean national controversy over the validity of a CPI. This last section focuses on elucidating what is implicated in effecting the invisibility of two different price-framing laminations.

The Aboutness of Price

In Chapter One, I described how Goffman (1974) defined frames as mechanisms for the production of aboutness. Goffman (1974) claimed that frames establish what situations are about; that is, they shape a situational sense of aboutness, and define the meaning of strips of activity. Frames serve to organise cognitive and bodily experience and they also provoke agential awareness of the meaning of situations. I would like to suggest that when the concept of framing is applied to data units like prices, one could also think of frames as mechanisms that participate in the production and organisation of *semiosis,* a term developed by Charles Sanders Peirce (1938-58) as part of his sign theory. In Peirce's theory, semiosis refers to the action of a sign, also defined as the *sign in process*. Semiosis is a dynamic and continuous triadic process that involves the establishment of relations and mediations amongst an object, a sign and an interpretant. For example, in one of Peirce's classic examples, a footprint in the sand is a sign, its object is the limb that created it, and its interpretant is the effect the sign created in the mind of the observer standing in front of the footprint. In short, semiosis is the effect that the relationship between footprint and limb might provoke in the human or non-human observer, for example the act of relating in thought: 'a human has recently walked in this spot' (see Delladelle 2000).

Semiosis was therefore for Peirce the relational process through which all three elements – object, sign and interpretant – were irreducibly brought together and interacted, a dynamic which was always generative of further semiosis. It could be argued then that any effects provoked by the irreducible triadic dynamism going on

between an object, a sign and an interpretant are instances of semiosis. It is worth noting however that in Peirce's theory, the sign, the object and the interpretant stand for relations and functions which are not fixed, but are changeable and multiple, to the extent that the interpretant in one semiosis can become the sign in a subsequent articulation, while the object of one sign could be the object of further signs, and so on. Peirce's theory of signs is therefore a theory of action, where 'the distinction between what refers and what is being referred to depends on the practical circumstances of the process of semiosis, and on its place on rather long chains of designation' (Muniesa 2007, 380). It could be argued that the practical circumstances that bound the process of semiosis are, as I have already described in previous Chapters, crucial if one takes into account Goffman's account of frames (1974). In other words, if one incorporates the notion of frame as developed by Goffman (1974) into Pierce's theory of signs it becomes clear that *semiosis* does not happen in a vacuum, but is organised and contextualised by frames and framing devices. These indicate when a triadic articulation between sign, object and interpretant becomes relevant and meaningful as about particular things and phenomena and not others; that is, they frame 'providences of meaning' (Schutz, cited in Goffman 1974).

Having suggested an incipient relation between Pierce's theory of the sign and Goffman's conceptualisation of frames, I will now turn briefly to describe a strand of economic sociology that has paid attention to the frames, devices and technologies that function to define the semiosis of price in particular ways. These studies recognise the situated, historically and technologically contingent formation of price, and have empirically tracked a variety of frames through which the aboutness of price has become configured as part of different market settings and practices (Muniesa

2007; Mackenzie and Millo 2003; Beunza and Stark 2012; Cochoy 2005, 2007a, 2007b, 2008). Fabian Muniesa (2007), for example, studied the techno-semiotic materiality of price closing mechanisms in the Paris stock exchange. He suggests that, as entities, prices are eclectic; they could be

> An agreement to close a transaction between two parties, a tag on a grocery shelf, an expert's indicative estimate of what a fair value of something should be or a synthetic input for a mathematical formula. Prices can be found or they can be proposed. They can be taken for granted or observed with expectation. They can be regarded as fair or unfair, as accurate or inaccurate, as good or bad, but also random, meaningless events. They can be thought of as coming from nowhere or as purposefully tinkered. They can be whispered in a conversation, handwritten in a personal statement or printed in a brochure. They can be talked about in various ways: I can 'give you' a price, but 'the market' can also give one. Prices are done, but they can do things too. Prices can be expressed in a quantitative fashion ('103.30 USD, 'Euribor plus 20 basis points') or in a rather qualitative one ('cheap'). (2007, 378)

In his deployment of Pierce's pragmatic theory of the sign, Muniesa (2007) draws out how prices come to refer to particular things or phenomena, highlighting the work required to establish such connections and purported aboutness. Muniesa (2007) notes how in financial settings such as the stock exchange, the closing price in a continuous trading session is important as it is often used as a reference for other calculations. Because of the importance of the price set for the last trade that closes the market, market agents became interested in manipulating such numbers in a way

that could be favourable to them. The contingency of price closure indicators was therefore expressed as a problem of representation: the closing prices at the Paris Bourse were deemed by market agents not to be valid representations of transactions in the market but expressions of particular interests instead. The problem was solved by the introduction of an algorithm called 'closure call auction' which prevented price closing manipulations and which technically settled 'closing prices'has more representative of genuine market operations instead.

Donald Mackenzie and Yuval Millo (2003) also argue that it was the intersection between the advances in pricing options theory, the further development of Black-Scholes pricing models, and the distributed deployment of computer technologies, that paved the way for objective price setting mechanisms, leading to the emergence of the financial derivatives market. Mackenzie and Millo (2003) demonstrate that as the Black-Scholes model was used more and more frequently by agents in the market, the actual prices in the market moved closer and closer to the ones predicted by the model – in other words, the model performed rather than anticipated the market. It could also be argued that the increasing use of models seemed to have direct effects not only on prices, but also on the interpretation attributed by market agents to the variation or fluctuation of those prices.

These two examples show how technologies and a range of other devices should be understood not only as agents that intervene in the formation of prices, but also how they contribute to an 'intersubjectively shared understanding of the "correct" calculation of prices' (Beckert 2011, 15). Devices like the closure call action algorithm and the Black-Scholes model have served to shape the collective cognitive

formatting through which agents interpret the situation they confront when observing market prices. In other words, devices like these ones work to diffuse and set a shared standard for the interpretation of prices, anchoring a frame of meaning for the collective reading of prices (Beckert 2011). Daniel Beunza and David Stark's (2012) work on the cognitive interdependencies established in financial markets through the widespread use of reflexive models is also explicit on this point. They look not at how price formation emerges through the use of models and other technical devices as Mackenzie and Millo (2003) do, but rather trace the effects that the use of models produce in the interpretation of price movements and noise, when the models work alongside scoping or observational devices (Knorr Cetina, cited in Beunza and Stark 2012). They argue that the interaction between models and observational instruments in financial markets not only aggregates otherwise dispersed information, but also distributes cognition, allowing market agents to read and think about what prices express in an interdependent manner. That is, these devices make cognition social. Such studies clearly show how devices and technologies enable the calibration of frames, which aids in the situated but still collectively distributed agreement and identification of what prices are about, what they express, and the kind of dispersed information they come to aggregate.

Whilst these accounts have looked at the stabilisation and framing of prices in the context of financial markets, others have explored the devices that contribute to shape the aboutness of prices in the more mundane settings of everyday consumer markets

instead.[77] Frank Cochoy (2005, 2007a, 2007b, 2008), for example, pays attention to how different material devices like shopping carts and packaging shape consumers' calculative and cognitive dispositions towards price. Cochoy (2008) claims, for example, that the use of the shopping cart as part of everyday shopping activities creates a storage zone that disassociates payment from choice. Devices like the shopping cart, therefore, enable consumers to put prices in 'parenthesis' (Cochoy 2008, 21). In other words, the shopping cart has come to facilitate the forgetting of prices in a literal sense 'since price labels remain stuck to the shelves'; but 'also figuratively as a few seconds after choice the consumer would not have the means or be equipped to evoke the values of the products chosen' (2008, 20). Cochoy (2008) also suggests that in supporting a volumetric as opposed to calculative purchasing capacity, the shopping cart equips consumers with calculative dispositions that incorporate qualitative assessments into the evaluation and comparison of consumer goods and their prices. This is a form of value assessment that has been defined as *qualculation* (Cochoy 2008; see also Callon and Law 2005; Callon and Muniesa 2005; Thrift 2004). The concept of *qualculation* therefore encompasses the idea that it takes material, semiotic and social effort to produce spatial practices, like the supermarket, that shape consumer cognition. The framing of the supermarket and its constitutive devices ensures that consumers establish particular relations with prices through which a form of price aboutness can emerge which is capable of bringing in qualitative elements.

[77] Although the differences in the operations and organisation of financial markets and consumer retail markets are striking, the conceptual framework deployed in STS to study the formation of price aboutness as part of both settings has so far been similar.

Cochoy's work has broadly addressed the functioning of different material and semiotic frames and techniques that configure physical spaces like supermarkets and retail outlets as particular types of markets, noting that the articulation of marketplaces, as with any other public arena, is a matter of '"air conditioning" and atmosphere management' (2007, 121; see also McCromack 2015). In tracing the work of frames and devices in mundane settings like the supermarket, Cochoy however also points out that windows into other frames can be opened up within the market frames of supermarkets, windows which in the cases he examined turn the consumer towards the 'outer world' (2007, 124). Cochoy presents one example that is relevant for the argument I will develop in the following sections of this Chapter. He notes that in 1999, under the political pressure of farmers voicing their concerns over the widening gap between producers and retailers prices, the French government issued a rule that stated that the production price and the retail price of some fruits and vegetables should be displayed across national retail outlets. Such disaggregated form of price had the effect of transforming 'individualistic consumers into consumer citizens able to evaluate the fairness of trading margins in the retailing sector' (2007, 123). Price decomposition became established as a new device through which to evaluate and qualify products in the context of consumer shopping. Although the rule lasted for only two months, it proved itself capable of transporting consumers – as shopping carts do – 'out of the realm of pure price economics' (2007, 124).

The studies described above have demonstrated how framing devices intervene in the formation of price and can elicit different forms of price aboutness, including, the forgetting, or bracketing, of prices. Furthermore, in the example of disaggregated prices that Cochoy (2007) provides, it becomes clear that multiple frames and devices

can operate and interact in the patterning of price aboutness. As Goffman (1974) argued, multiple frames can be in operation for the same strip of activity; alternatively, as in this case, different frame laminations and layers might come to configure the situated aboutness of a singular price at a particular point in space and time. The material and semiotic framing of the retail outlet and its internally operative devices might frame and thereby stabilise what a price is about, but the pattern of such framing can also be used to establish other types of 'price aboutnesses' too. For Goffman (1974), all frames have a recursive nature; this is made evident by the fact that in Goffman's conceptualisation, not only can all frames be reframed, but reframed frames can then be embedded or recontained as part of more (complex) re-framings. In addition, a given frame's reframing can be made visible either totally or partially. If the latter, then some of the actors might not be aware of the reframing taking place, although they might be physically contained or bounded as part of such material and semiotic reconfiguring. Goffman (1974) defined these recursions as *frame laminations,* and in the following sections I describe the devices by which such different price framing laminations might become visibly superimposed on each other or not.

Product Identification

It is how laminations of frames articulate different modes of price aboutness that I would like to further engage with here. As Cochoy argues (2007, 2008), consumers are affected by the material and semiotic framing of price experience performed by retail outlets and supermarkets. More recently, Johan Hagberg and Jans Kjellberg

(2015) have studied how the configuration of price as part of different physical formats organises particular shopping and pricing experiences, focusing on material forms such as labels, signs, posters, bundles and brochures which are often displayed alongside product descriptions. Price differentiation, comparison and temporalisation are techniques for organising consumers' everyday sense of market prices, and such framings of price aboutness are usually made visible to all consumers or participants brought together as actors in the market.

However, further price reframings take place that are not made visible for consumers to assess, sense or experience. The identification and measurement of the products and prices taken into account for the calculation of Consumer Price Indexes (CPI), for example, take place in the same settings but are not made discernible to consumers or citizens. Generally, and for reasons I explain in the section below, consumers are prevented from knowing which particular prices and products are being taken into account within shopping settings. This partial invisibility occurs however despite the fact that the measurement of CPIs constitutes a further lamination, or reframing, of the aboutness of prices in retail settings that consumers inhabit. The measurement of CPIs, I claim, imposes a *frame lamination* that produces a contextual setting patterned on but different to the one retail outlets provide. CPIs use the pattern already laid down in the framing accomplished by consumer retail markets, but they stabilise the aboutness of price differently by re-configuring the relations that prices and their contexts establish with goods, services and consumers.

Let me here first briefly and in very broad terms describe what a CPI is. A widely accepted definition of a Consumer Price Index is that it is an indicator of the average

measure of change in the prices of a set basket of consumer goods and services bought for the purposes of consumption by a representative average of households of a specific geographical territory (Consumer Price Index Manual, International Labour Office 2006). This territory is sometimes, but not always, the territory of a nation state; in the case I will discuss, for example, the CPI related only to a particular region of Argentina – the Capital Federal and Greater Buenos Aires region. As such the geographical referent of this CPI is not the country as a whole. Nevertheless, the territory to which it refers is an administrative region of the Argentine nation, and it is the Argentine state that has responsibility, and a large administrative apparatus, for the production of this indicator on an ongoing (monthly) basis.

CPIs are examples of composite numbers (Guyer 2010); that is, the CPI signifies – in a single number - an ensemble of heterogeneous things that exist in reality in a differentiated and distributed manner (Desrosières 1998). Moreover, CPIs are measures that transform a complex phenomenon into a single value that varies over time (Bruno et al 2014). CPIs are also generally designed as 'pure price indexes': 'pure' here signifying that only *changes in prices* between different reference periods should be expressed by the index. This process of purification requires the isolation of price factors over a fixed (base) period of time, a comparison with the same factors in a second period, and then an estimation of the change that occurs if everything else - such as consumers and their spending patterns, the market and the goods and services it offers, the quality of the goods and services on offer, the condition in which products are encountered and so on - is taken out of the equation. Price factors are only 'those which identify the amount paid for one unit of a specified kind and quality of good or service obtained from a specified source by a specific population

group' (Hurwitz 1962, 815). This process of purification attempts to ensure that any variation in price observed is not contaminated by other kinds and modes of variation. Indeed, it is the task of the CPI methodology to render these other modes of variation demonstrably *external* to any observable variation in price.

Significantly, purifying price variation entails among other things that the products which have their prices observed and annotated as part of the index remain qualitatively constant and uniquely identifiable in space and time. This ensures that the recorded variation in price is attributed to changes in the pure price of 'a same thing' over time, and that those changes do not express changes in the quality or quantity of the products on offer. The identification of goods and services to be measured as part of CPIs is usually informed by a classification system[78] that is internally differentiated into divisions (e.g. Food and Non-Alcoholic Beverages), groups (Food), and classes (Bread and Cereals). Using this classification system, the task of a national state CPI methodology is to produce and publicise a number that represents the variation in price of a weighted average of representative products and

[78] The classification structure of most Consumer Price Indexes (CPIs) is based on an international classification system developed by the United Nations Statistical Division called named COICOP, Classification of Individual Consumption According to Purpose (COICOP). It is called a 'functional' classification because it identifies objects in terms of the 'function' of consumer transactions. The unit of classification that emerges is 'expenditures on goods and services';: that is, COICOP groups together goods and services that are deemed to fulfil particular consumption purposes, such as 'nourishing the body, protecting the body against inclement weather, preventing and curing illness, acquiring knowledge, travelling from one place to another' (ILO Consumer Price Index Manual, Theory and Practice, 2006, 59).

retail outlets, the publication of which is fixed at an agreed level of classification or aggregation.[79]

To go about the process of observing the price of a class such as Bread for example, a significant assumption is made, namely that the prices of different kinds of Bread vary in a similar fashion. This allows the price collectors to justify selecting only a sample of items. For example, the classification system might go as far as specifying that 'large white loaves – sliced and unsliced', 'large wholemeal loaf', 'bread rolls' and 'garlic bread' might be selected as products that, if they vary in price, can legitimately be taken as representative of the price variations for all instances of the class of Bread. But this classification is not in itself sufficient for the identification and observation of the price of Bread, as bread does not present itself as 'a large, white loaf, sliced and unsliced', as its retail ontology is always configured as more diverse and specific. The layering or lamination of the CPI's classification system on top of the market frames that form part of the operation of retail spaces therefore performs a gap between what are conceived as generic forms of Bread and their diverse enactments.

[79] Some other countries provide more disaggregated data relating to products that are specifically identified more closely specified. So, for example, the UK Office of National Statistics (ONS) recently released what is defined as 'CPI's micro-data', following a Freedom of Information Act request that was filed in 2011 demanding the publication of individual item indices. Since then, ONS has extended the network of attributes that make visible each of CPI's specified products as part of its Open Data agenda. However, because of other regulatory frameworks still in place, products are often still unidentifiable in their actualised forms, in large part because brand names are still excluded from public data releases in the UK.

But whilst CPI methodologies problematise the gap that they establish between classes of abstract products and the diverse specific enactments of such products in different retail outlets, they also include a range of techniques that address this difference. This is not only important because matching abstract classes of products and 'actual' products is required to measure price variation, involving a particular mode or way of identifying products (as I shall describe below). It is also important because the process of matching and identification according to CPI methodologies takes advantage of the frames that retail outlets organise, in particular with regards to how these settings articulate and display the qualification of products available. The qualification of products as part of market settings where consumer goods are exchanged, like supermarkets or retail outlets, assign and display properties that help to objectify and singularise goods (Callon and Muniesa 2003). According to Callon and Muniesa (2003), two crucial steps are generally involved that guarantee the singularisation of products: firstly, a product needs to be 'profiled', which consists in 'establishing a space in which it can be connected and compared to finite list of other products and assigned with particular characteristics and properties' (2003, 6). Secondly, singularising a product also entails linking the product to other similar products in a process of 'classification, clustering and sorting' that makes products comparable but at the same time different to each other (2003, 14). Goods can therefore be calculated or qualculated, and all these operations 'constitute the material base for the extraction of a result (a price, a classification, a choice)' (2003, 150).

CPI methodologies rely on such processes of qualification of products as part of their price measurement. As part of the methodologies of CPIs, products are typically further specified in ways that allow them to be identified reliably both in a specific

space and time and – potentially at least - in other spaces and times for subsequent measurements. The way in which this is achieved in practice is by reference to the further specified qualities of the products that consumer retail markets configure, such as: the characteristics of the goods as such ('white' bread); aspects of the way products have been produced; a product's packaging characteristics; their distribution pathway or country of origin; and in some cases, a product's brand name too. Such use of these specified qualities of the products expressed in consumer retail markets not only serves to breach the gap between classification classes and particular products, but also ensures that the products for which prices are being measured remain constant over time. This in turn certifies that prices therefore express pureness of variation, as they are made to purposefully fluctuate according to CPIs methodologies.

This process of specification is standardised and reproduced each time price measurement takes place as part of CPI methodologies. Figure 2.0 in the Appendix shows the process of specification and identification of products that price collectors are trained to carry out when entering a consumer retail outlet where CPI measurements are to be taken. On arriving at a retail outlet, price collectors need to find an already specified product – a specification that I have shown is achieved combining an abstract classification of goods with the qualities of the products that retail outlets have on display. If the specified product is not available at that point in time, price collectors need to find a similar variety of such product, record their qualities in a sufficiently specified way so as to determine any differences in quality and provide a description that will enable an exact identification of the variety in future measurements.

The process outlined in the diagram and which I have briefly described above can be conceived of as a technique for stabilising the formation and aboutness of price. As Muniesa (2007) points out, the establishment of an algorithm – in the case of CPIs a rule-based way of matching and identifying products, as shown in the flowchart – ensures that the correspondence of prices to products is of a particular 'pure' kind, and that prices are made to signal nothing more than variation. The establishment of this particular correspondence between prices and products thus requires a considerable amount work and, in the case of CPIs, it is accomplished through the use of algorithmic techniques of product identification and descriptive specification in the field. Whilst it could be argued that in many situations prices serve to qualify products in that they are temporarily made 'to attribute a set of characteristics to an entity for the purposes of exchange' (Hagberg and Kjellberg 2015, 184; Callon and Muniesa 2003) CPI techniques for the stabilisation of price aboutness work the opposite way. They make the specification and description of products qualify prices as pure instead.

So far I have described two different frames that configure the aboutness of price, and the types of relations that prices establish with a number of other good and entities as part of everyday consumer retail markets. On the one hand, there are the framings and devices that articulate the relations that prices establish with consumers as part of everyday shopping practices and activities, and which enable different forms of price aboutness. On the other, I have also described how the patterns of those settings are redefined when the price reframing instigated by CPIs methodologies becomes active. These two framings of prices operate sometimes simultaneously, but the CPI's

frame very rarely becomes fully visible for the actors engaged in everyday shopping activities. In the following section I engage with the activation of one legal device that has forced the patterns of these two frame laminations to become - at least simultaneously - indiscernible from each other. In my exploration of this device I argue that making the workings of these two frames visible would necessitate a further reframing, one through which the aboutness of price could become again reconfigured.

Secrets and Data Aggregation

In the previous section I have described how technologies, devices and frames contribute to stabilise the aboutness of price. The stabilisation of the aboutness of price usually also entails the production and display of price as a singular - if composite - number; or rather, as a whole and fully-formed number. Jane Guyer (2010) however suggests that there is a new moral economy of composite numbers emerging in society at large. Whilst composition - that is, adding parts up to give a single price - is usually not recognisable in the customary representation of prices as singular and text-based numbers, Guyer (2010) claims that contemporary practices of price decomposition should be understood as *a method for revealing prices' moral and political constitution*. As the example offered by Cochoy (2007) above shows - where the prices of fruits and vegetables became disaggregated to express retailer and producers' margins throughout consumer retail outlets in France – these types of practices have served to establish relations between previously unrelated registers or frames. In other words, they can be thought of as openings of frames within frames.

Guyer (2010) illustrates this point by showing how in Nigeria it became politically imperative for oil prices to express their logic of composition, and in so doing they explicitly traced and connected the oilfield to the petrol pump as part of everyday settings. Guyer (2010) argues that in the case of Nigeria, the national decomposition of the price of oil into different identifiable parts 'gave rhetorical traction on the political reality of price' (Loc. 5396) - a reality that would not have been intelligible to consumers and citizens if posed in technical or purely economic terms. In what follows I trace the attempted decomposition of a national CPI database in order to explore how 'disaggregation' could have reconfigured the aboutness of price, and in what ways.

Let me start by briefly describing such attempted decomposition or disaggregation of CPI. On the 5th of February 2007, the Institute of National Statistics and Census in Argentina (INDEC) released a press statement:

> The General Level of the Consumer Price Index for Capital Federal and the corresponding districts of Greater Buenos Aires registered in January a variation of 1.1% in relation to the previous month, and of 9.7% with respect to the same month in the previous year. (Pedido de Indagatoria 2007, 99)

Since then, this regional Consumer Price Index for Capital Federal and Greater Buenos Aires (CPI-GBA) and its subsequent variations have been at the centre of a national and international political, legal, and methodological controversy. The legitimacy of the numerical value of the percentage has been called into question by a range of actors and has been – as I will show in Chapter Six - challenged by a variety

of both public and private alternative measures since then. Throughout 2007 in Argentina, a series of public denunciations[80] of the numbers presented as the CPI-GBA led to the launch of an independent legal investigation, carried out by the Fiscalía de Investigaciones Administrativas (FIA)[81] in order to determine whether the CPI-GBA had been manipulated in an inappropriate way, and to ascertain whether it made adequate, representative and reliable reference to the fluctuation of prices in the region. The public denunciations were initially mostly concerned with challenging the government's intervention[82] in the activities of INDEC, but they also questioned the accuracy and epistemological validity of the CPI-GBA produced by INDEC (see Figure 2.1 in the Appendix). A range of actors alleged that the normal procedures for the production of historical series of the CPI-GBA had been altered without introducing a corresponding clear-cut change in the methodological guidelines for the

---

80 The initial denunciations were made by statisticians working at the INDEC at the time of the intervention, denunciations that were covered in all the main newspapers in the country (La Nación, Clarín, Perfil and Página 12). It is worth noting that the Asociación de Trabajadores del Estado (ATE), a branch of the Central de Trabajadores de Argentina (CTA) was particularly supportive of such denunciations. The FIA received two requests as part of the investigation: the first one was filed by the Asociación Civil Sin Fines de Lucro 'Asamblea por los Derechos Sociales' led in 2006 by Dr. Aníbal I. Faccendini. The second one was filed on the 14th February 2007 by a number of national senators.

81 The FIA is an independent organ of government dedicated to the investigation of corruption and administrative irregularities undertaken by agents within the national administration of government. The FIA mainly undertakes preliminary investigations, and facilitates the filing of legal cases.

82 In its most literal sense, the intervention refers to the moment in the controversy in which long-standing statisticians were dismissed from the INDEC and substituted by others who were, more closely aligned with the government's stance. However, the intervention may also be understood as a moment in which the legitimacy of the terms of the aboutness of price as an indicator was brought into question.

collection, aggregation and weighting of prices.[83] Demonstrators also declared that primary data had been deleted and replaced with secondary data, that did not form part of the measuring procedures and sampling techniques traditionally used in the production of the CPI-GBA at the time of the controversy.[84] In addition, it was further questioned whether the data was produced in accordance with the terms of the legal framework designed to preserve national statistics, and whether the data abide to the peculiar relation these terms are meant to establish with the settings, prices and products in and through which they become calculated.

I would like here to focus only on this last set of concerns, specifically the queries relating to Law Number 17.622 of 1968, article 10 and 13, concerning the need for *statistical secrets*. The legal documents which emerged from FIA's enquiry, and the interviews I conducted with statisticians and technicians dismissed from INDEC, suggest that members of the government demanded that INDEC provide a detailed disaggregation of the prices and objects of measures taken into account to construct CPI-GBA. This included a request to the professional statisticians employed at INDEC to provide details of the specific retail outlets and brands of products being taken into account in the production of the CPI-GBA.

---

[83] For example, it was alleged by interviewees that the database that was used to produce the Index was altered so as to allow – inappropriately - the limits on price change agreed by the government and some companies with some companies in relation to some categories of goods - to inform the calculation of the CPI.

[84] My analysis of legal materials and the interviews I conducted indeed suggest that the secondary data collected and produced by the Secretary of Commerce, and not that collected by INDEC, was included as part of the production of CPI-GBA for January 2007. See Berumen and Becker (2011) for further discussion.

On the 11th of July 2006, Guillermo Moreno, then Secretary of Commerce, made the first of a number of formally documented requests to INDEC, asking for the release of data which would reveal not just the 'Variety' of products –for example 'Oranges' - but the further specification of the products selected to produce CPI-GBA. He also requested the details and geographical coordinates of the retail outlets at which the prices of such specified products were being observed. Government officials justified the request by saying that the release of the information was necessary so that the government could face 'the serious problem that inflation represented for the country, manage the inflation expectations of the population and the impact these might have over the future economic development of the nation' (Guillermo Moreno – Secretary of Commerce, Pedido de Indagatoria 2007, 1; my translation).[85] Despite repeated requests, the INDEC statisticians refused to disclose information to this degree of specification to the government.

Until the so-called intervention, INDEC had routinely published changes in prices to the level of a product's Variety; that is, it made average price variation public to the level of a relatively unspecific generalisation, such as for example, Oranges. When questioned at the legal enquiry, one of the INDEC statisticians responded to questions regarding the request by the government for more specific information by saying, 'if we had released information of this kind [that is, at the level of specific products in

[85] The request for this information by the government was widely interpreted to be part of an attempt to manipulate the CPI-GBA. While some commentators also argued that such requests were motivated by the desire to correct the perceived plutocratic bias reproduced in the Index, others suggest that the government's intention was actually to decrease the value of debt bonds linked to the CER (Reference Stabilisation Coefficient), which involves the use of the CPI-GBA.

specific places] to Moreno, we would have handed him the elements to regulate and control the prices of the specified products and that would have had an effect on the Index' (Pedido de Indagatoria, 29-30; my translation). It was claimed that the disclosure and decomposition of data to such a degree could destabilise the frame that CPI routinely deployed to measure pure price variation. The following hypothetical account was offered as a means of justifying the workings of the Statistical Secret Act in relation to CPI-GBA:

> A government official finds out that the INDEC only measures the price of 1 litre of milk in the sachet packaging of Brand X to account for the variation in the price of all milk in general. For statisticians, this random sample is enough to believe that other specifications of the same product move in a similar trend, that is, that the price of a single milk sachet can represent that of all milk. The government official [however] can seek to agree a price with milk producers for that product in particular, and in exchange for such a favour, he can authorise price increases in all other milk and dairy products which do not form part of CPI's goods to be measured. If such was the case, in the final press release, a decrease in the price of milk would have been registered, although in reality the price of milk would have increased. (Juegen and Bullrich 2010, 28)

The reason given by the statisticians for their refusal to give the information was thus the desire to prevent government intervention in the production of the CPI at the level of the manipulation of specific prices of specific products. In other words, the statisticians were concerned about the performative effects that the release of

statistical data would provoke. To justify their refusal, they referred to Law Number 17.622 which sets out the terms of the regulatory framework which requires that both the specific characteristics of data, and the combinatory possibilities that may be established between different data units, be kept secret. As discussed in the introduction, a Statistical Committee was also convened to determine and review the rules that governed the disclosure of such information. The Committee determined that the data could only be released as long as such release would not enable the public identification of the products and locations used for the calculation of CPI-GBA.

As I have noted before, the formation of CPI prices necessitates that particular links are established between these prices and the qualification of consumer goods that retail settings configure. But this brief description of the CPI-GBA controversy shows that the formation of price not only establishes and sustains particular links with products, but also creates and maintains a whole range of disconnections. In refusing the release of micro or detailed data, the statisticians demonstrated their belief that a 'sphere of obscurity and a structure of withholding' is required for the legitimacy of the CPI-GBA: their actions were directed towards protecting a space of secrecy from which 'the data could first be developed and from which the figures could emerge as part of its process of entering the public sphere' (Didier 2005, 639).[86] The figures then were designed to enter the public sphere in a composite or aggregated manner - that is, without making the locations and products used for the production of this number identifiable. This also meant that the patterning of the convergent frames of

[86] That is, it sought to protect the details of their first order observation from any second order observation other than their own (Esposito 2013; Stark 2013).

retail outlets and CPI measures were also protected from further reframings, and obscured from public view.

Indeed, the refusal by the statisticians to release CPI in a disaggregated manner points to the importance of the legal restriction on such possible reframing, not simply as a barrier to manipulation, but as a device that can be activated and operated by the statisticians themselves. Such a device is one that blurs the visibility of the multiple frames in operation that shape the aboutness of price as part of everyday consumer settings. Such a device also enables data to become framed so that actual products may be identified by the statisticians and changes in their price can be observed as part of their calculation of the CPI, while simultaneously detaching the prices of those products from their socio-technical, everyday and political environments for everyone else. In this way, the operation of the legal restriction understood as a device allows for prices to be made available for statistical observation and measurement based on the already framed context of the retail outlet and its geographical location, whilst making such context and the specificities it configures unobservable to everyone else than the statisticians.

I want to suggest here that the disaggregation of data - which would have identified the settings, products and prices that were taken into account to calculate CPI - would have entailed a subsequent reframing of the aboutness of price. Such reframing would have accomplished the *contexting* (Asdal and Moser 2012) of price more extensively to incorporate and make visible other parts that might participate in the formation of price. Such reframing would have further contextualised the prices of products, revealing other potential factors that contribute to price aboutness. Such an

experiment in contexting[87] (Asdal and Moser 2012) would have in the first place opened up the aboutness of price to include more elements of the patterned frames from which it had been decontextualised. But in the case described in particular, decomposing CPI's data would have constituted a reframing that would have revealed, for example, how the variation in the price of *generic* Oranges differs from the variation in the price of Organic Navel Oranges sold in the gentrified neighborhoods of Capital Federal, which in turn differs from the price of residual Osage Oranges sold in the popular markets of some Greater Buenos Aires districts. That is, it would have enabled socio-demographic data to participate in the aboutness of price (Uprichard et al. 2009).

Indeed, some commentators have suggested that the reason the government requested information from the statisticians was to correct the plutocratic bias[88] that CPI-GBA was deemed to suffer from as part of its calculus; that is, the government's request for this information was seen by some to be motivated by a desire to show that the specification of products being selected by INDEC for the calculation of the Index was not representative of the consumption patterns of the population as a whole.[89] But if the data about which *specific* Oranges had been observed for changes in their price was *not* disaggregated, then not only would the government or other agencies be

[87] In this sense, instances of data disaggregation could be understood as an experiment in contexting (Asdal and Moser 2012). Following the tradition of interactionism and ethnomethodology, Asdal and Moser (2012) have defined experiments in contexting as activities that might rendermake the contextual effects that they themselves unleash visible.

[88] In an interview conducted with an adviser to the government on statistical matters, it was claimed that the intention of the data disaggregation 'exercise' was to correct the plutocratic bias that the CPI-GBA carried.

[89] See for example www.indec.com.ar, last accessed on 16th April 2014.

prevented from manipulating prices, but also the contexting of price in relation to specific localities or specific income groups would be made indiscernible. This containment of the overflow of data however would not prevent the public CPI composite figures from intervening in the everyday consumer settings from which the figures initially emerged, as economists have already demonstrated how consumer inflation expectations are influenced by the taking into account of publicly available information such as CPI composites (see for example Carrillo and Erman 2009; Cavallo et al. 2014). The claim being made was that the disaggregation of data would have enabled a more extended contexting of price.

I have described here the working of a legal device that prevents the patterning of the frames that intervene in price formation as part of everyday life from becoming publicly observable. It is a device that organises and contains the extent to which price might relate to certain things or phenomena. In this case, the withholding of data - as required by law, and upheld by the statisticians - sustained the processes by which the measurement of price is protected from manipulation at the level of the observation of specific products. But importantly, it also allows statisticians to render price as pure, referring simultaneously to the real and to nothing but itself (in a previous time period). My argument is that the operation of this legal device prevents a type of price aboutness that includes elements that lie beyond the ones provided by the retail settings in which prices became annotated as part of CPI-GBA. In other words, (legal) secrets prevent a mode of contextualization defined as 'out-contextualisation' (Lury 1998), entailing the multiplication of contexts that promotes a framing of an object, in this case prices, as if they could be seen from all positions at once.

Conclusion

This Chapter has examined the emergence of price aboutness as part of the workings of different framing techniques and devices. I have firstly suggested that one way of conceptualising the aboutness of price is to combine a Peircian pragmatics of meaning with Goffman's take on frames as processes which contextualise semiosis, in order to frame the triadic and irreducible relation between a sign, an object and an interpretant. I have then described how Peirce's theory has been used in the work of Muniesa (2007) to account for the formation and interpretation of price, and how the work of other economic sociologists can also be understood as descriptions of the framing devices that enable prices to become observed and interpreted in socially interdependent ways. The Chapter has paid particular attention to the framing of prices in consumer retail settings that provoke a form of cognition defined as 'qualculation', arguing that supermarkets and other types of retail environments frame the experience and sensing of prices in a multiplicity of ways. I have argued that these are not the only frames operative as part of consumer retail spaces but that subsequent frame laminations and rekeyings of price can occur, with or without the knowledge of (some of) the participants interacting with and in that space. Goffman (1974) understood reframings as laminations that occur over but also within an already established frame, involving the transformation of something which is already meaningful into something that actors should evaluate as different but still patterned on a previously articulated frame.

In this Chapter I have also described the identification techniques that are deployed in the production of the CPI in order to illustrate how retail prices participate in a variety of frame laminations, and how these techniques make use of the framing of price found in retail outlets by rekeying it to account for price variation in time. I have suggested that whilst the frame of consumer retail outlets and the frame of the CPI operate for the same products and prices, the superimposition of both frames as two different modes of organising the aboutnesss of price is rendered invisible to consumers. The containment of a leak of from a national CPI database which could have made the superimposition of frames visible shows that such activity would have enabled a further reframing of price to include socio-demographic components as part of its aboutness. Moreover, the legal containment of price aboutness prevented a mode of contextualisation which would have been enabled by the leak of data, defined as out-contextualisation (Lury 1998).

Chapter Six

Digital Methods and The Financialisation of Retail Price

Introduction

In the previous Chapter, I accounted for a range of framing techniques and devices that configure the aboutness of price in offline settings such as supermarkets. This Chapter follows a similar set of concerns with regards to the framing of prices, but this time with a particular focus on techniques and devices that have emerged as part of the framing of *digital* prices instead. Arguably, these devices and techniques form part of a newly emerging set of digital methods for the collection and recording of data (Rogers 2013), used both commercially and academically for different motives and objectives. The aim of this Chapter is to investigate the challenges that these digital techniques and devices pose for the formation and aboutness of price, in order to understand how framings of digital price might intervene in how the economy and its fluctuations becomes observed and interpreted.

In his essay entitled *The Use of Knowledge in Society,* Friederich Hayek (1945) claimed that observations of the state of the economy and its fluctuations are misleading if they focus on the whole economic picture provided by aggregate statistics, instead of on the movements of the units which comprise them. Hayek (1945) argued that the knowledge that social scientists should be preoccupied with is that which cannot enter into statistics, and 'therefore cannot be conveyed to any

central authority in statistical form' (1945, 524). Central or official statistics, Hayek (1945) claimed, are arrived at by 'abstracting from minor differences between things, by lumping together, as resources of one kind, items which differ as regards to location, quality, and other particulars', and that these differences 'could be very significant for the making of economic decisions' (1945, 524). The use and communication of official statistics, he argued, do not take into account particular circumstances of time and place that affect the decision-making processes of the 'man on the spot' (1945, 525). The 'man on the spot' however, cannot make economic decisions solely based on his subjective and intimate knowledge of the facts that populate his immediate surroundings (Hayek 1945). A system is required for accumulating and then conveying information about 'the whole pattern of changes of the larger economic system' to the man in his particular spot (Hayek 1945). Such a system, Hayek (1945) claimed, is configured by *the price mechanism*. He further argued:

> It is more than a metaphor to describe the price system as a kind of machinery for registering change, or a system of telecommunications which enables individual producers to watch merely the movement of a few pointers, as an engineer might watch the hands of a few dials, in order to adjust their activities to changes of which they may never know more than is reflected in the price movement. (1945, 527)

He went on to note in relation to the coordination function enabled by the price system that

> The marvel is that in a case like that of a scarcity of one raw material, without an order being issued, without more than perhaps a handful of people knowing the cause, tens of thousands of people whose identity could not be ascertained by months of investigation, are made to use the material or its products more sparingly; i.e., they move in the right direction. (Hayek 1945, 527)

Hayek (1945) distrusted the knowledge content of economic statistics, aggregate analysis and averages because, he argued, these 'tend to obscure micro-level dynamics and give a misleading impression of greater stability in relationships over time than what in reality exists' (Bronk 2012, 89). Hayek's vision of the power of prices as signals that serve to coordinate the economy emerged from a context in which aggregation and statistical composition were not, as he originally thought, the outcome of an epistemic imperative but more the product of methodological and technical constraints. Both governments and the man on the spot needed to collect and asses information that was dispersed throughout the economy, and to have access to signals which could express in a unified way any economic change that happened in a distributed manner across space. However, aggregating, condensing and composing data in the form of statistics were for Hayek (1945) mechanisms for reducing complex, dynamic and distributed phenomena into a singular centralised figure that did not intervene in the coordination of social functions - which did not capture the function of prices as far as Hayek (1945) was concerned.

Things however have changed since Hayek's prominent essay on the use of prices for economic decision-making, and I argue that the way things have changed in

particular challenges the critique Hayek made in relation to centralised uses of knowledge in the form of official statistical aggregations. Firstly, the tension between the amplification gained by transforming local economic knowledge into aggregated forms and the reduction effected in the loss of locality and particularity is becoming more tenuous, as instruments, technologies and devices make it possible to move from the local to the general and back again - what is described in the work of Latour (1999, 2010), and throughout the thesis, as *traceability*. Traceability undermines Hayek's argument about centralised ways of assessing and observing the economy, as it is now possible to dis- and re- aggregate data in order to make localities and particularities visible without debunking the function and usability of the aggregate. On the other hand, the digitalisation of more and different registers of economic life, and the production and circulation of statistics based on these new registers, mean that economic observations (and hence economic decisions) are now also necessarily more varied. Latour (2010) offers a glimpse of what the proliferation of measures and statistics might provoke when elaborating on Gabriel Tarde's notion of quantification:

> If statistics continues to progress as it has done for several years, if the information which it gives us continues to gain in accuracy, in dispatch, in bulk, and in regularity, a time may come when upon the accomplishment of every social event a figure will at once issue forth automatically, so to speak, to take its place on the statistical registers that will be continuously communicated to the public and spread abroad pictorially by the daily press. Then, at every step, at every glance cast upon poster or newspaper, we shall be assailed, as it were, with statistical facts, with precise and condensed

> knowledge of all the peculiarities of actual social conditions, of commercial gains or losses, of the rise or falling off of certain political parties, of the progress or decay of a certain doctrine, etc., in exactly the same way as we are assailed when we open our eyes by the vibrations of the ether which tell us of the approach or withdrawal of such and such a so-called body and of many other things of a similar nature (Tarde, cited in Latour 2010, 115-116)

Such a prediction resonates uncannily with contemporary understandings of the information society. These changes challenge the function of economic statistics and open up spaces to rethink how they should be calculated, communicated and used. It is important to note in this respect that at the time I am writing this (July 2015), the UK government has commissioned a review of official economic indicators in the country. The review will consider whether the current producers of official statistics are well-placed to keep pace with the latest advances in data collection, particularly in the light of a myriad of commercial digital data initiatives that have emerged to take advantage of newly digitalised data types that feed into economic decision making processes (see below). It is believed that as 'tech firms learn ever more ways of extracting and combining information, the UK's premier source of economic data resembles a rusting and poorly organized filing cabinet.'[90] Moreover, the emergence of so-called Big Data might pose a challenge not only to official statistics but also to economics as a discipline in itself. Although the term Big Data as such has not yet gained any significant traction in economics, 'the availability of unprecedentedly rich datasets and the need for new approaches – both epistemological and computational –

[90] Lies, Damned Lies and a Better Path for Statistics, *Financial Times*, Retrieved from www.ft.com, last accessed 31st August 2015.

to deal with them is an emerging issue for the discipline' (Taylor et al. 2014).

In order to account for these changes and what they might imply for the observation of the economy and the coordination of economic action, the following Chapter focuses on a particular innovation that was made possible by the attempted leak of CPI data and the controversy this generated as described in Chapter Five. The controversy around the validity of CPI opened up a space for the emergence of alternative indicators of inflation. Among these indicators was the Billion Price Project, a Big Data experiment that started as an academic exercise but later developed into a commercial exercise, and subsequently informed the commercialisation of alternative measures of inflation for a range of different purposes, including the assessment of retail price fluctuations in financial settings. In tracing and analysing this particular intersection between official public numbers, alternative measures and the commercialisation of public data, I attempt to examine and compare the different data framing devices and techniques deployed for the production of offline versus online (and digital) inflation measures, in order to understand what this might imply for the framed organisation of price aboutness. The first technique I study is that of online data *scraping* that was developed as part of the Billion Price project. The second device I study is the *image,* understood as a mechanism of data decomposition developed by a venture data enterprise called Premise. In both instances, I study how these techniques and devices frame the aboutness of price and delimit the type of relations that prices might bring about and establish with a range of objects and phenomena in the world. I do so in order to account for the implications that the digitalisation and commercialisation of economic data have for the ways in which prices and their fluctuations might become observed

and interpreted.

Scraping

In the previous Chapter, I described how the method of price collection employed as part of CPI calculations serves to limit the aboutness of price, and how it relies on but also further patterns the framing articulated by retail outlets. I have also shown how devices operate to make the entanglement of these two frames publicly indiscernible. In this section, I look at a technique for the collection of online digital prices called *scraping.* I claim that this technique relies - as the measurement of offline CPI does - on the framing of prices established by digital retail formatting, and that scraping also repurposes these in order to further reconfigure the aboutness of price. The patterning and rekeying of frames for the repurposing of data like prices is therefore something that occurs as part of both digital and analogue measurement practices and methods. In other words, it could be argued that practices and technologies of data annotation and extraction have always entailed aligning, overlaying and patterning different frames. However, differences between digital and analogue practices lie in how the context and aboutness of price is accomplished and observed, and what types of frame visibilities digital methods like the ones described in this Chapter enable.

Because the emergence of digital methods for the measurement of inflation has been closely linked to the controversy around the offline CPI in Argentina, I would here like to start by briefly contextualising this particular empirical relation. Following the publication of the findings of the investigation carried out by the FIA as described in

Chapter Five, the controversy over the validity of the CPI-GBA continued to grow, and many other national indicators ended up being challenged - including, for example, measures of poverty and extreme poverty, real earnings, real exchange rate, GDP, and so on (Lindenboim 2013; Becker 2013; Berumen and Becker 2011). Nevertheless, some provinces continued to estimate their own regional CPIs to release to the public,[91] and some of these numbers came to be regarded as the 'real' measures of inflation in the country. At the same time, rival measures to the CPI-GBA were also developed by academic economists and representatives of various private organisations,[92] who developed independent, non-governmental projects to observe the prices of a basket of goods and services in order to estimate inflation rates. These typically produced a figure for variation in prices that was significantly higher than the official calculations produced by INDEC for Capital Federal and Greater Buenos Aires.

It is worth noting however that the emergence of alternative CPI indicators is not unique to the controversy that unfolded in Argentina, nor to the possibilities offered by digitalised data collection methods. A study by Alain Desrosières (2014) for example traces the history of alternative price indices from the beginning of the

[91] Dirección Provincial de Estadística y Censos, Provincia de San Luis (DPEyC – San Luis); Dirección Provincial de Estadísticas e Investigaciones Económicas, Provincia de Mendoza (DEIE – Mendoza); Dirección General de Estadística y Censos, Provincia de Córdoba (DGEC – Córdoba); Instituto Provincial de Estadística y Censos Provincia de Santa Fé (IPEC – Santa Fé); Dirección General de Estadística y Censos, Ciudad Autónoma de Buenos Aires (DGEC – Capital Federal).

[92] Buenos Aires City; Gabriel Rubinstein & Asociados; Estudio Bein & Asociados; EconViews; M&S Consultores; Ecolatina; FIEL; Abeceb.com; Asociación de Dirigentes de Empresas (ADE); Finosport; and Orlando Ferreres y Asociados, among others.

1970s to the 1990s in France. Throughout this period, the General Confederation of Labour (CGT) in France published the results of an index that it had built itself, in opposition to that produced by the National Institute of Statistics and Economic Studies (INSEE). The CGT argued that the INSEE index was based on assumptions that corresponded to the consumption patterns of the middle classes rather than that of the working class, and produced an alternative figure that was deemed sufficiently trustworthy to be used in national wage settlements. The disclosure of CPI that this involved however still only disaggregated data to reveal a certain general or generic level of specification of products.

In Argentina, the response of the government to these alternative measures was to ban them, and fine the organisations producing them on the grounds that they were misinforming the consumer. In 2011 the Secretary of Commerce, Guillermo Moreno, fined private economic consultancies on the basis of on an alleged violation of article 9 of Law Number 22.802. This law stipulates that any public presentation, publicity or propaganda containing inexactitudes leading to consumer deception or confusion may be sanctioned. The government considered that the publication of alternative inflation indexes that lacked scientific rigour could induce confusion in members of the public, leading them to make ill-informed decisions, and that such publications should therefore be banned on the basis of this law.

Indeed, the government suggested that the only entity with the methodological capacity to assess the CPI rigorously at the geographical level of either the Capital Federal and Greater Buenos Aires or the nation was INDEC; they asserted that such a complex, large-scale operation was economically and logistically impossible for

private organisations. In response to the ban by the government, the private consultancies came together in June 2011 under the banner of IPC Congreso, an initiative through which members of the opposition (including some of those belonging to the Commission for Freedom of Expression in the Cámara de Diputados) made the figures produced by these private consultancies public in the Congress, in order to avoid being fined. However in October 2013, the Supreme Court ruled that the government fines were illegal and private consultancies began to publish their inflation figures independently again. More recently still (January 2014), the government released a newly designed national CPI (IPCNu) that (implicitly at least) addressed the technical and political flaws identified at the time of INDEC's intervention. In its first release the new indicator, which has been welcomed by a range of stakeholders - including the media, economic experts and international bodies - has resulted in an inflation figure that is widely accepted as a more accurate expression of the real fluctuation of prices as encountered in the consumer retail economy.

I want here to look at one particular spin-off measure that emerged in response to the controversy, that also attempted to provide an alternative inflation figure: the Billion Prices Project (BPP; http://bpp.mit.edu/, last accessed on 15th June 2015). This project provided a challenge not simply to the particular figure that was put forward as Argentina's CPI, but also to the production of the CPI as a national statistic more generally, since it involved the use of online or so-called 'big data' rather than the offline data obtained by statisticians as described in the previous Chapter. The project was initially created by Alberto Cavallo and Roberto Rigobon, two economists, now based at the MIT Sloan School of Management in the USA. BPP currently monitors

the daily price fluctuations of approximately 5 million items sold by approximately 300 online retailers in more than 70 countries, in order to produce, among other things, a daily price index and inflation rate for a number of countries, including Argentina. The BPP originated in 2007 as part of Cavallo's PhD thesis[93] at Harvard University, which compared the online price variations for Argentina, Chile, Brazil and Colombia. Cavallo later went on to produce an alternative online inflation index for Argentina called 'True Inflation' which was accessible via a webpage,[94] and updated price variations for Argentina on a daily basis. In 2008, data collection was expanded to 50 other countries, an exercise which officially became the BPP. What started as an academic exercise, however, was later subsumed in a company that describes itself as 'the leading source of daily inflation statistics around the world', trading under the registered trademark of PriceStats.[95] Indeed, the BPP website now refers those seeking 'more high-frequency inflation data across countries and sectors' to PriceStats, 'the company that collects the online data we use in our research initiatives and experimental indexes'.[96]

As indicated above, what is interesting about the methodology of the BPP is that it made extensive use of online price data, taken from the largest supermarket chains, which it collected using scraping software. As a method, scraping is substantially different to the procedures used to specify and find products off-line in order to observe price fluctuation. While the production of CPIs using established statistical methodology requires the repeated, and constantly varying involvement of human

[93] See Cavallo (2010).

[94] See http://www.inflacionverdadera.com/, last accessed 11th June 2013.

[95] See http://www.pricestats.com/, last accessed 29th October 2013.

[96] http://bpp.mit.edu, last accessed 31st August 2015.

agents (price collectors) and devices (questionnaires, flowcharts, instructions, etc.), scraping performs a similar but a much more automated set of operations over a relatively frictionless surface produced by HyperText Markup Language (HTML).[97] Cavallo (2010) describes scraping as follows:

> To collect the online scraped data, I programmed a software that scans the underlying code of public webpages, finds relevant price information, and records it in a database. This software can be set to run automatically every day, visiting a retailer's website, accessing the links to pages with product data, and collecting detailed information for individual goods on display. The resulting data contain daily prices for all products sold by the retailer over time. This information can be collected at a relatively low cost in any country with online retailers. Furthermore, scraped prices can be used to build datasets with identical products, time periods, and sampling characteristics, making it easier to identify robust stylized facts across retailers and countries. (2010, 3)

BPP's scraping technique was further described by Cavallo as a three-step process. First, selected web-pages which display product and price information are downloaded. Second, the underlying code of those web-pages is analysed by software to locate each piece of relevant information to be scraped: 'this is done by using custom characters in the code that identify the start and end of each variable,

[97] As per Cavallo: 'The primary language used to write content on the Web is called HyperText Markup Language, or HTML. It is written using tags, such as < *center* >, which provide instructions for the browser to render a page graphically to the user. These tags can also be can be used by the scraping software to identify relevant product and price information in the code.' (2010, 3).

according to the format of that particular page and supermarket' (Cavallo 2010, 7). Third, scraped information is then stored in a database containing 'one record per product per day' (Cavallo 2010, 7). The nature of the data collected and the possibilities offered for the formation of price aboutness through the method of scraping differs substantially to that conventionally used in the production of CPI as a national statistic. Cavallo claims that

> Scraped data have important advantages that make them a unique source of information. First, these datasets contain daily prices, which can greatly reduce measurement error biases (…). Second, the data are available for a much larger set of countries (…) Third, scraped data contain detailed information on the full array of a retailer's products. In particular, the ability to identify products displayed next to each other plays a key role in measuring price synchronization among close substitutes. Fifth, there are no forced item substitutions, which occur frequently in official statistics to measure inflation in out-of-stock, seasonal or discontinued products. Sixth, scraped datasets are directly comparable across countries, with prices on the same categories of goods and time periods. This makes it possible to perform simultaneous cross-country analyses. Finally, scraped data are available on a real-time basis, without any delays to access the information. This can be used to provide estimates of stickiness that quickly capture changes in the underlying economic conditions. (2010, 7)

So while scraping was created in part to confirm the belief that the official offline inflation figures in Argentina were inaccurate,[98] BPP is far from being just a substitute methodology for the production of CPI. Indeed, the aboutness of price is more variously organised through scraping than through the production of national statistics like CPI. A whole series of contrasts can be drawn between BPP and CPI in terms of how they configure the aboutness and formation of price differently. Firstly, although the BPP follows traditional CPI classifications and respects official weightings of prices for the purposes of producing a measure that is comparable to conventional CPIs, scraping invites the experimental exploitation of price through the use of multiple (and possibly changing) kinds of categorisations, thus allowing different kinds of price formations and comparisons to be made across these multiple categories. In addition, the time periods over which price can be demonstrated to vary in relation to any and all of these categories can be intensively differentiated, allowing for both very short-term and long-term comparisons.[99] Scraped data enables many different kinds of calculative operations, providing the basis for prices to become framed and be put in many different kinds of relations to other entities and phenomena too: so, for example, BPP produces measures of real-time global inflation; price synchronisation among brands and contiguous products; price stickiness; price mark-ups, etc.

---

[98] Cavallo notes that 'amongst their many potential uses, price indexes constructed with online data can be used to obtain alternative inflation estimates in countries where official estimates have lost their credibility. In particular, this paper uses online prices to evaluate the widespread claim that the Argentine government has been manipulating the official inflation index since 2007' (Cavallo 2012, 1).

[99] Although arguably scraping also poises problems of discontinuities too, see Boyd and Crawford (2012); Marres and Weltevrede (2013).

As can be seen, scraping expands and extends the contexting of price. Whereas the CPI secures one form of price aboutness – that of pureness of variation – BPP on the other hand enables prices to be contextualised in a much more undetermined way. Such fluidity in contexting is what grants the data – in this case prices – to be of economic value to a range of different actors. While arguably the publication of CPI has socially and historically regulated the economy of economic observations, by digitally scraping prices and enabling them to be put in relation to a range of other data units to confirm a series of different contexts, the observation of retail price fluctuation is becoming economised - that is, accruing its own market value (Caliskan and Callon 2009, 2010).

Let me here however return to another important difference between CPI and BPP. I want to turn briefly here to think about the functions of CPI and inflation measures within society at large. Derek McCormack (2015) has recently suggested that price stability is central to the security of specific valued forms of life in the functioning of western liberal democracies. Threats to price stability, McCormack claims, have been 'named and governed as emergencies', and governments are held responsible for containing and managing such events (2015, 141). McCormack proposes a genealogical understanding of how the relation between scarcity of goods has been historically configured as a problem of security. He draws attention to how the relation between price and security has come to figure quite prominently in the emergence of governmental neoliberal thinking during the 20th century, claiming that

> Instead of something left to vary as part of the natural functioning of the market, price – or price aggregates – is now something that needs to be stabilised in order that economic life and market activity can take place. In this context, the "main objective of regulatory action will necessarily be price stability, understood not as fixed prices but as control of inflation" (Foucault 2010: 138-139). As he writes, what is therefore to be secured or saved, "first of all and above all, is the stability of prices" (ibid, 139). Critically, as Foucault makes clear, in this model of government, price is not something upon which one should act. But we might therefore ask precisely what it is that efforts to govern price act upon if they do not act upon price? (McCormack 2015, 134).

McCormack (2015) examines these efforts to govern price alongside the wider question of how price emergencies are handled and how the stability of prices becomes managed and articulated in varied forms – as part, for example, of popular culture practices too.[100] He notes the ways in which inflation has figured in 'governmental concerns and everyday life as an emergency in and of itself, one that

[100] McCormack for example claims in relation to the TV show *The Price is Right*: '[prices become governed] semiotically, and affectively, a structure of feeling that is embodied in consumer experiences such as The Price is Right, the longest running game show in the US, which first aired in 1956. As a consumer experience the show obviously celebrates the importance and spectacle of price in everyday life. It also demonstrates vividly, however, how prices are participants in the generation of affective spacetimes. Affects – hope, disappointment, despair – gather round prices. And The Price is Right deliberately engineers and modifies the affective atmospheres that gather around prices, things, and bodies. In this context, what counts as a good price is not just one that is right in terms of the process of veridiction about which Foucault writes. Insofar as a price is right, it is one that feels right. And this feeling is distributed across bodies'. (2015, 135).

demands and justifies particular kinds of juridico-political, technical and practical responses' (2015, 135).

It could be argued, following McCormack's thesis, that the production and making public of a legitimate and objective CPI is a technique for socially governing price emergencies. Contrary to how Hayek (1945) envisaged centralised statistics as not playing a role in organising the economy, one of the functions of accomplished public and objective measures like CPI is, I argue, to organise how the public observe the economy, make inferences, and speculate about its present and future outlook. Indeed, one of the remarkable functions of public statistics and other forms of common knowledge is that they are made visible to all agents so that they become established as a shared reference, an observation about the state of the economy that is not only available to everyone, but more importantly that 'everyone knows to be known' (Esposito 2013, 14). This is because, whilst CPI is accomplished as an epistemically valid index of a given phenomenon such as inflation, once made public also it also works as an index of what others observe (Esposito 2013).

Sociologists have argued that that by mediating and coordinating the mutual observation of observers, numbers and measures modulate the formation of expectations amongst agents as part of different settings (Esposito 2013; Espeland and Sauder 2007). This, I argue, also holds true for official statistics and numbers. Public and hence symmetrical information, that is, information equally available and distributed to all agents at the same time, has a social value and function because it intervenes and serves to coordinate how agents observe and relate to each other, whilst modulating the inferences that agents make of each other's observations in

relation to the state of the national economy (Didier 2005). The ways in which statistics become publicly disclosed matter in this respect, as I have demonstrated with the example of CPI in the previous Chapter, because the degree of aggregation ensures that a frame becomes established and secured for the collective agreement of what the numbers are about and how they fluctuate in time.

In information economics, the social function of public numbers and prices in particular has also been recognised in a number of different ways. Paul Carillo and Sehan Emran make the case, for example, that 'governments provide public information about market conditions partially in order to reduce information imperfections and facilitate efficient allocation decisions in the economy' (2010, 1). Economic agents use a range of information sources to inform their decisions, information that also contributes to shaping agents' expectations about future economic activity. Carillo and Emran experimentally demonstrate that public signals like CPIs play 'a crucial causal role in the formation of households' price expectations' (2010, 2). Because public signals - such as published official statistics - aggregate dispersed information across the economy, they reveal information about the actions of others, demonstrating the importance of public indicators like CPIs in the formation of expectations and their latent role in potentially configuring heard behaviour and social learning among other phenomena (see also Cavallo et al. 2014). In other words, the publicity of CPIs serves to organise the social observation and experience of prices, and in so doing might play a crucial role in the governance of latent or actual price destabilisation crises or emergencies too. As I have shown, such organisation of the social observation of price variation normally works only if the CPI number is not entirely decomposed, that is, if the number is made public to a

level of aggregation that does not render the specificities of particular price variation patterns visible.

It is interesting in this respect to observe that the data 'scraped' and distributed by PriceStats is made available to others in commercial, academic as well as public forms, each defined by varying terms of access. Thus the PriceStats website says that it 'distributes' its daily inflation statistics[101] through an exclusive partnership with State Street Global Markets,[102] whose target clients include hedge funds, pension funds, and sovereign wealth funds. It also states that it is in partnership with the BPP, described by PriceStats as 'an academic initiative that uses high frequency price information to conduct breakthrough economic research'; and that it collaborates with public institutions to improve decisions in public policy: for example, they create special indices that measure the price of specific goods across countries to 'anticipate the impact of commodity shocks on low income, vulnerable populations'.[103]

In this regard it appears as if BPP and its commercialisation and publicisaiton through PriceStats has the potential to be used not only to challenge the epistemic validity of the CPI as a national statistic in particular countries or nation states like Argentina, but also to offer new opportunities for the economisation of price variation in that it establishes different frames for the observation of price aboutness for different groups of actors. It enables the financialisation of inflation insofar as it provides data so that economic actors can put multiple categories of price variation in relation with a range

[101] That is, presumably, it sells this information.

[102] See http://www.statestreetglobalmarkets.com, last accessed on 16th April 2014.

[103] http://www.pricestats.com, last accessed 19th September 2015.

of different phenomena. This makes it possible not only to detach the change in prices from any normative or exclusive relation to the State's observation of price variation, but also from wages, pensions, savings, and the reproduction of a population or cost of living in any particular country. In this respect, such framing of price variation organises a different type of economic observation through which – as the next section suggests – price emergencies may be exploited rather than governed. This is, paradoxically, a new plane of what Isabelle Bruno et al. (2014) have defined as 'statactivism', in this case afforded by the production of macro-economic indicators produced as Big Data numbers and facilitated by particular methods like scraping. It is paradoxical because the contestation of statistics is being undertaken for economic rather than libertarian purposes.

Imaging

As I have shown in Chapter Five, some of the frames that shape the aboutness of price variation in everyday life are indiscernible. The legal device described in the previous Chapter worked by obstructing the disaggregation of data to the level at which it would reveal, as part of everyday shopping and settings, the prices and products taken into account for the measurement of CPI in Argentina. In the following section however, I would like to concentrate on a recently developed technique of data disaggregation that has been developed commercially and that works the opposite way: it enables, via the collection and distribution of pictures and maps, the identification of the locations and products where inflation prices are being recorded and measured. This technique of data disaggregation undertaken via images

has been developed by Premise, a recently established privately-funded start-up[104] which tracks local and global inflation trends. This new agency describes itself as follows:

> We founded Premise in 2012 because existing mechanisms of econometric data gathering are insufficient to meet the policy, trading, relief or business strategy challenges posed by an era of unprecedented economic and social volatility. Our experience at the grocery store had nothing in common with the inflation headlines we were reading in the paper. We wondered, what happens when you replicate and scale this divergence across ever-larger groups of people? Headlines over recent years call into question just what is going on with the global economy, with exports, with employment figures, with interest and currency rates… We grew tired of feeling like no one really knew what was happening.[105]

Premise describes its mission as the creation of 'faster, more accurate, more liquid data sets to deliver the critical transparency our electronic age demands', and as a way of 'unearthing connections that impact global decisions'.[106] Its methodology was described in one newspaper report as 'a blend of Google Street View and the CPI',

---

[104] See https://www.premise.com, last accessed 31st August 2015.

[105] See http://blog.premise.com/news/2013/10/15/introducing-premise/, last accessed 29th June 2015.

[106] See Premise, http://www.premise.com/introducing_premise.html, last accessed 29th October 2013.

or, alternatively, as 'an international photo-collage project',[107] and involves the collection of proprietary price variation data from a number of what are defined as 'unique sources', including e-commerce websites, third-party transaction streams and their own mobile workforce which visits 'real-world' stores in Argentina, Brazil, China, India and the United States to 'ground truth directly' with the use of a mobile app that allows employees to photograph products and prices with a time and date stamp. In its most recent promotional materials, Premise suggests that 'a picture is worth a thousand data points' and that the company brings 'hyperlocal visibility to the world's hardest-to-see places' by 'blending technology with human intelligence' in order to map (this time) 'reality to the ground'.[108]

Using and blending data from different sources, Premise initially offered a range of different products including, at the time of writing (October 2014), a global Food Staple Price Index alongside Individual Food Staple Price Indices for a number of countries, including Argentina, USA, Brazil and India. It also commercialised indices for 'sub-components' of interest such as Meat, Fruit, and Vegetables, defined as Food Security Indices. More recently, Premise started to diversify its offer to include not only alternative indicators of inflation for different countries, but also bespoke experiments to estimate real-time price stability and variation, incorporating variables like weather phenomena, queuing times, stock availability and product placements. It has also recently moved away from being a purely price monitoring venture to one which also commercialises other data types, for example, the identification via

[107] See http://www.theatlantic.com/business/archive/2013/10/the-price network/280553/, last accessed 14th April 2014.

[108] See https://www.premise.com/#home, last accessed 15th June 2015.

satellite imaging of un-electrified areas in East Africa aimed at companies intending to invest in electricity infrastructure.[109] This follows a trend in the exploitation of satellite micro-data for economic purposes, as shown for example in the commercial developments of Orbital Insight Inc., a company that has recently commercialised data sets of 'changing shadows of Chinese buildings' which according to the company can provide 'a glimpse into whether the country's construction boom is speeding up or slowing down'.[110] Orbital Insight analyses satellite images of construction sites in 20 Chinese cities 'with the goal of giving traders independent data so they don't need to rely on government statistics'.[111] Other examples of Orbital's products include the analysis of satellite imagery of cornfields to predict crop yields, and studies of satellite imagery of commercial parking lots to provide early indicators of retail sales and quarterly earnings.[112]

Particularly with regards to inflation measurements and experiments, Premise's network of contributors take pictures of products and consumer items, including their positioning on the shelf or in a market stall, which automatically sync with Premise's servers. Premise's analytic systems are then able to extract a range of different information from the picture such as price, brand and quality of the items, and even

---

[109] See https://www.premise.com/case/#1, last accessed 31st August 2015.

[110] Hope (2014, November 9), Startups Mine Market-Moving Data From Fields, Parking Lots – Even Shadows, *Wall Street Journal*, retrieved from www.wsj.com, last accessed 28th November 20014.

[111] Gannes (2014, October 13), Premise Creates a Global Economic Activity Tracker, One Snapshot of Onions at a Time, *Wall Street Journal*, retrieved from www.wsj.com, last accessed 31st August 2015.

[112] See Kurgan (2013) for a compelling critical and political account of the use of satellite images and technologies in different fields and as part of everyday life.

contextual information such as how clean the store is and how stocked the shelves are. The pictures of each of the products taken into account, and the different range of inflation Indexes that Premise delivers, are offered to clients as a source of insight and added economic value. Let me here draw on one of Premise's advertised case studies to elaborate on how this modality of disaggregation might serve to take advantage of rather than govern price emergencies. This also offers an interesting point of comparison with the social and public CPI functions outlined above.

Premise drew on its identification of the prices, quality and stocks of Onion in India as a way of demonstrating how its data could help investors, business leaders and decision makers forecast economic movements. In an interview, Premise's founder, David Soloff, claimed that official economic indicators are too outdated by the time they are released and that although people 'lurch to the number', the world has moved on since these indicators were calculated and published.[113] The scenario that Premise used to illustrate this point was one in which, despite unprecedented price spikes in onions during the second half of 2013, the RBI (the Indian Central Bank) only responded to an increase in inflation with a change in interest rates in the last quarter. According to Premise however, the company detected a significant increase in price per kilogram of onions as early as June 2013, as reported by the company's street teams who were making daily records of 7,000 different onion prices in markets and stalls in cities such as Mumbai, Chennai and Kolkata. The street teams' pictures also showed an unhealthy change in onion colour, a diminution in size and a

[113]Gannes (2014, October 13), Premise Creates a Global Economic Activity Tracker, One Snapshot of Onions at a Time, *Wall Street Journal*, retrieved from www.wsj.com,; last accesed 31st August 2015.

reduction in stock size, all changes which Premise interpreted as an indication of shortages and hoarding (See Figure 3.0 in the Appendix). 'We see the peak being talked about when it's already being resolved' Soloff said. 'We have a reliable granular read on inputs, rather than waiting for official bodies to speak about it. We've built a better mousetrap to capture data at scale'. [114]

The picture frame therefore offers a way of contextualising price differently. It can relate price variations to phenomena like emergencies or price spikes, relations which traditional CPI figures do not disclose in their composite arrangements. So whilst the disclosure of compositional and aggregated data in the case of the CPI described in the previous Chapter secured a frame for the collective observation and understanding of price variation, in the case of Premise the decomposition and disaggregation of data is what enables its clients to observe price trends - as emergencies and therefore opportunities - that are unavailable or indiscernible for others to see. The disclosure techniques deployed to make official statistics like CPI public also dictate that this data is available according to a temporal – usually quarterly and retrospective – sequence. This is to ensure that noise reduction in the informational content of these measures is reduced and to establish temporal release points that permit equality of access to the figures by members of the public - although there are exceptions, as pre-release access can also be granted (see Didier 2005). However, it is the granularity of the data that Premise delivers, combined with

[114] Gannes (2014, October 13), Premise Creates a Global Economic Activity Tracker, One Snapshot of Onions at a Time, *Wall Street Journal*, retrieved from www.wsj.com,; last accesed 31st August 2015.

'nowcasting'[115] temporalities, which makes Premise's products economically valuable to its clients. Here disaggregation too can be understood in a rather Hayekian (1945) fashion, as the most important knowledge with respect to economic activity is not the knowledge of centralised statistical measures but that of particular circumstances of time and place. This type of knowledge - highly dispersed, fragmented and local – is usually rolled up into public statistical summaries like official CPIs. In this respect, Premise presents the detection (and eventual governance) of inflation emergencies as a *local knowledge problem*, that is, it configures the data for the detection (and exploitation) of price emergencies as existing in a disaggregated manner.

What the onion example in India serves to demonstrate is that Premise's methods configure and exploit emergencies as well as providing a pre-emptive observational frame of retail price variation for those willing to pay for it. This is a form of data collection that distinctively seeks to align itself with some of the mechanisms deployed for the collection of state data like CPI, but in order to provide instead 'emerging and unfolding futures' (Amoore 2011, 27). As Louise Amoore (2011) argues, data derivatives like the ones commercialised by Premise are not in the 'same order of being as what we might call modernist disciplinary data' (2011, 27). Rather,

[115] In economics, the use of the term *nowcasting* refers to the design and use of statistical models based on real-time or close to real-time data that offer short-term forecasts of different conditions and phenomena including the state of the economy but also the weather, epidemiology, humanitarian crises, political opinions and national sentiments among others. The term is a contraction of the words 'now' and 'forecasting' that is meant to reflect the capacity of statistical models to offer a picture or reading of the temporal dimensions of the present, the near future and the recent past.

they come into being through the amalgamation of previously unrelated data – in this case prices, products and their qualities, and shelves, as a means of identifying stock availability and anticipating looting behaviour. To elucidate this new kind of data, let me introduce Esposito's (2012) descirpiton of the use of derivatives as financial instruments

> Derivatives are, in fact, so named because they have the particularity of "deriving" from something else. Their price is calculated from the price of something else, the "underlying", which can be anything – for example, natural facts like the amount of snowfall in a resort, the wheat harvest, the price of pigs or other financial instruments like stocks, indexes, rates. (2012, 105)

Following Amoore (2011), I suggest that the data produced by companies such as Premise might be called data derivatives, and as such stand in sharp contrast to the type of data generated and aggregated by official statistical formulations. I have shown how CPI can be thought of as an element that contains and governs how populations observe and experience price variation. The techniques which generate and re-present the data sold by Premise, however, are not geared towards inscribing what prices are at a given point in time, or even in real time. Rather, they frame prices in terms of their proclivities and potentialities – in terms of what they *could be* (Amoore 2011). This is done, as I have shown, through the framing of prices alongside the 'anytime anywhere emergence of the abnormal' (Massumi, cited in Amoore 2011, 32), such as when the methods deployed by Premise enable associating price spikes with the colour of onions.

Whilst the previous Chapter showed how the patterned frames of retail outlets and the measuring of CPI were made publicly indiscernible as a way of stabilising the collective aboutness of price variation, the introduction of measurement procedures examined in this Chapter, like the one offered by Premise, makes these framings visible as a new form of economic value. The data, in the form of pictures being nowcasted or released in advance of official CPI figures, facilitates the private and commercial observation of possible price emergencies. Here the configuration of decomposition and granularity of data, alongside nowcasting temporalities and the extended context of price, serve to financialise and securitise retail prices; that is to configure the framing of retail price for financial purposes. The financialisation of retail prices entails the exploitation and not only the containment of price emergencies. In Chapter Five I described two frames that are already operative for defining what prices are about in retail settings: the framing of supermarkets that enable different calculative consumer operations to unfold, and the framing of CPI which secures a particular form of price pureness. To these frames we should now add the one organised by the calculative operations of Premise. The superimposition of these frames, and the devices and practices which enable these superimpositions to become visible to some agents and invisible to others, are crucial in the context of the commercialisation of retail prices. This is because it is the partial but advantageous view of the overlapping lamination of such frames that endows prices as data with economic and political value.

Conclusion

In this Chapter I have looked at the digital devices and techniques that frame prices in the calculation of alternative digital inflation measures, and suggested that they challenge the functioning of traditional economic statistics by enabling an expanded contexting of price variation on the one hand, and the reframing of inflation as a local knowledge problem on the other. The first section of the Chapter described the adoption of a digital technique of data collection defined as *scraping,* used for the measurement of inflation, in the wake of the controversy generated around the CPI in Argentina. Whilst the collection of digital prices also implies reframing HyperText Markup Language to account for price variation, and in this respect resembles offline measuring inflation procedures, I have claimed that scraping differs from offline procedures for framing price in a number of ways. Firstly, through scraping, the aboutness of price can be more variously organised by experimentally exploiting price in the deployment of multiple categorisations. Moreover, scraping also enables a different temporality of price variation by allowing for short-term and long-term price fluctuation comparisons. Also, by scraping prices, they can be put in relation to other entities and phenomena and their contexts expanded. I have thus claimed that in contrast to CPI, methods like scraping enable prices to become contextualised and their aboutness stabilised in a much more undetermined way, a function which in turn secures the economisation of digital prices.

This Chapter has also investigated the social function of inflation measures like CPI. I have claimed that the publication of a legitimate and supposedly objective and public inflation measure can be understood as one of the mechanisms available for governing price emergencies. Whereas Hayek (1954) downplayed the role of

aggregated statistical measures in the organisation of social structures for the observation of the economy, I have argued that public measures like CPI should be understood as playing a crucial role in the organisation of the social observation of the economy and its fluctuations. This is because public statistics can act as observational coordinates: a shared reference of the state of the economy amongst a population of observers. In other words, public measures function as 'indexes of what others observe' (Esposito 2013, 16). Because they constitute a shared frame of reference, public statistics modulate the formation of expectations in the population with regards to how prices might vary in the future. The social function of measures like CPI might be changing, I argue, because of the emergence of alternative indicators which are becoming commercialised and which elicit alternative social and private observations of inflation. Initiatives like BPP establish different frames and perspectival views of price aboutness, and because these are not made public, they organise a different social mode of observing the economy only for those who pay for its access.

The final section of the Chapter has proposed that these new digital price framings are organised not to establish a common frame of observational coordination, but as a means of financially exploiting variation in price. I have proposed that the digital image or picture can be understood as a frame that enables price to emerge as a local knowledge problem. This in turn enables it to frame prices in terms of their proclivities or potentialities. This was exemplified with the example given of onions in India by Premise. The detection of the change in colour of onions, present in the pictures taken by Premise's crowd-sourced teams working across India, became a way of framing the possible price and stock emergency of this particular staple in the

country. I have claimed that the image framing of decomposed price variability, alongside the use of nowcasting temporalities and the extended contexting of price which relates it to more and different phenomena and entities, can be understood as techniques and devices that serve to financialise retail price. This is because these framings enable the exploitation rather than the containment of price emergencies as a way of anticipating eventual emergencies. In financialising retail price, what appears fundamental is the partial view that these proprietary frames enable. The view is limited because the frames are proprietary, and hence they configure and organise a mode of partial observation for anticipating rather than for collectively assessing the variation of price.

Case Study II - Price Formations and Statistical Compositions

Conclusion

Whereas in the first case study I looked into the making and unmaking of personal data, this case study has focused on studying the making and unmaking of prices as economic data. This case study has examined a range of techniques and devices for the framing of price. I have described how through the framing operations of these devices and techniques, prices become formed and the functions of statistics and data defined. This case study has shown how frames are important in shaping prices in economically significant and bounded ways. This is not only because frames can render prices economically valuable as data as I have shown in Chapter 6, but also because the ways in which frames bound prices configure economic phenomena like inflation and shape economic cognition, contributing in this way to organise economic action too. I have based my analysis on the theoretical grounding of Pierce (1931-58) and Goffman (1974), and on the conceptual developments in economic sociology and the social studies of finance that have paid attention to how frames, devices and technologies function to both define and unleash the semiosis of price as part of different settings, from stock exchange floors and international securities offices to retail outlets and supermarkets. I have adhered to a similar range of conceptual propositions advanced as part of this literature, but have focused instead on revealing the importance of the lamination, superimposition and limited visibility of the multiple frames that are in operation to frame price as part of retail settings and inflation measurements. I have also paid particular attention to how methodological procedures and legal devices are implicated in articulating the semiosis of price,

whilst acknowledging some of the differences between field and digital methodological techniques regarding the framing and formation of price.

A range of conceptual themes have emerged as part of this case study which I would like to briefly dwell on in this conclusion. The first theme relates to the lamination of frames, or the conceptual possibility offered by conceiving multiple frames as being operative at the same time, as opposed to thinking about a singular frame defining the aboutness of a situation, event or space - in this case data. For Goffman (1974), different frames can be in operation for organising the experience of a situation, and as I have described in Chapter One, frames can be rekeyed but they can also be fabricated. The difference between a frame *rekeying* and a frame *fabrication* is that in the former agents are organised to interpret the frame in a given way, that is, the frame is visible and understood by all participating agents; whereas in the latter frames are hidden from view for only some agents to interpret. The notion of frame fabrication was developed by Goffman (1974) to take account of situations where the agents included within a frame are not aware of either the existence of the frame or the different rekeyings that are taking place to further laminate the frame they are, or have unintentionally become, a part of. Goffman (1974) identified two types of fabrications: on one hand, benign fabrications are those that are intended to organise experience 'for the benefit of those it deceives' (Manning 1992, 126). On the other hand, exploitative fabrications constitute framings that are deployed and orchestrated for the 'benefit of the fabricator' (Manning 1992, 126).

This case study has engaged with the superimposition of frames and the implications that the visibility, transparency or opacity of these laminations pose. I have shown

how frames (and the information they carry) can be concealed from view or made publicly available for all agents to see. The concealing and disclosure of information is a topic that has been widely studied sociologically. Georg Simmel (1906), for example, attributed information concealment – such as the configuration and distribution of secrets in society –with particular effects and functions (see also Galison 2004). Simmel suggested that 'if there were such a thing as a complete reciprocal transparency, the relationships between human beings to each other would be modified in a quite unimaginable fashion' (1906, 448). Simmel affirmed that the existence of degrees of intensity and shading in which each individual revealed herself to others was a characteristic of modern social relations. Such calibration and 'credit economy' established between revelation and secrecy among human actors is, Simmel argued, one of the principal methods that sustains interpersonal relations in society. In Simmel's account of secrets, the reciprocal knowledge exchange, and more importantly the knowledge of reciprocal concealment, underpinned the formation of modern social relations. Simmel claimed that 'the secret of the one party is to a certain extent recognized by the other, and the intentionally or unintentionally concealed is intentionally or unintentionally respected', and he therefore hailed secrecy in this respect as 'one of the greatest accomplishments of humanity', as with 'publicity many sort of purposes could never arrive at realization' (1906, 462).

Secrecy was, according to Simmel, useful and functional in the organisation of human reciprocities and economies of social interaction. Simmel described secrecy as a 'sociological technique' and as 'a form of commerce' without which certain social accomplishments could not take place (1906, 468). Both Goffman's account of fabrications and Simmel's legacy on secrets are useful to think about how

information concealment and disclosure on one hand, and the visibility or invisibility of frames on the other, can structure and organise particular modes of observing and interpreting both social and economic phenomena. The structures and ways of organising observation and interpretation can be of different kinds, and produce a wide range of effects: they can bring about symmetrical and collective ways of observing and interpreting phenomena; or they can enable individual and asymmetrical modes of observation and interpretation, to name a few combinatory possibilities.

Through this case study one could further argue that the superimposition of frames that define what prices are about in space and time on the one hand, and the different perspectival views that are a consequence of such superimposition on the other, should also be thought of as *functional* in organising how the economy and its fluctuations is observed and interpreted, and how agents respond to it. In this case study, I have shown that the devices and techniques being deployed to frame prices – be these offline, in the field, or digital – organise the observation and interpretation of the economy and its fluctuations asymmetrically. In the case of national statistics, such asymmetry is organised by the limits imposed on the public disaggregation of inflation data and the impossibility of visualising, as part of retail spaces, which are the consumer goods taken into account to measure inflation. This asymmetrical misinformation of consumers, producers and government officials as to which products account for inflation is functional. It is functional in this case because it enables the measurement of pure price variation whilst reducing measurement reactivity, that is, it reduces the possibility of agents changing their behaviour due to the incorporation of this type of information into their economic - and political, as the

Argentinian CPI controversy shows - calculations. If the products and prices used to calculate inflation are publicly communicated, both through the release of disaggregated data or through their display and signage as part of retail outlets, a different organisation of experience with regards to the stability or instability of prices would be configured, having a possible effect on the formation of inflation expectations by the public in turn (see Braun 2015).

In this case study I have also shown how the framing of economic fluctuations through the scraping and imaging of retail prices permits a financialised interpretation of price movements. I have also compared how the disaggregated and dispersed framing of prices elicits a different mode of observing and interpreting the economy compared to the interpretations made possible by the publication of national statistics in an aggregated manner. I want to suggest that whilst the function of the latter can be understood as being like an *observational coordinate and focal point*, in the sense that public statistics are references of what all agents in the economy know to be known (Esposito 2013), the privatisation of inflation by Big Data commercial experiments and enterprises might challenge the symmetrical coordination of the observation and interpretation of the economy and its movements that public statistics have afforded so far. This is because private Big Data analytics enable a dis- and re-aggregation of units, and they frame data for the production of local knowledge. They also modulate the temporality of the release of data differently whilst commercialising such temporal differential. This framing of privatised inflation data thus produces an asymmetrical and individualistic observation of economic variation, through which, as I have suggested, price emergencies may be exploited rather than governed.

Conclusion

Data Types and Functions

This thesis has explored the framing of data. I have started in Chapter One by elaborating a conceptual framework that could account for the *framing* and *contexting* of data, and that could shed light on the different methods, devices and techniques being used to make data about different entities and phenomena in the world. I have firstly looked at the lineage of the concept of *frame,* and how it has been developed in cybernetic anthropology, cognitive frame theory and social interactionism as a way of challenging taken-for-granted understandings of contexts as self-evident and given. I have paid particular attention to the notion of frame as found in Goffman's work (1974). I have described how, in Goffman's account, frames can be understood as mechanisms for the production of *aboutness*, that is, as mechanisms that define situations as being about different phenomena and things. I have also shown how frames for Goffman serve to establish a sense of aboutness because they organise experiences that elicit an awareness of the meaning of a situation or strip of activity.

I have noted the prominence of the notion of *laminations* in Goffman's theorisation about frames. I have described how in Goffman's terms, multiple frames - either superimposed, layered or juxtaposed with each other - can operate at a given time for a given strip of activity. I have accounted for how Goffman classified these laminations as *rekeyings* or *fabrications* depending on the degree of visibility of different frames in any given interaction, and the amount of information that is available to the actors participating in those framings. Also in Chapter One, I have

described how Callon (1998) extended the Goffmanian notion of frame further by advancing the concepts of *device* and *overflow*. I have demonstrated how the notion of device is helpful to think about non-human actors as crucial participants in framing operations, and how we should not only pay attention to the containment of strips of activities by frames, but also to the destabilisation and instability of frames. The first Chapter turned on Callon's (1998) argument that the different elements that form part of the frame represent openings onto other frames. There are in this respect no static, enclosed insides to frames, but open configurations that make possible the fleeting stabilisation of certain connections.

In Chapter Two, I have described the research design pursued by the thesis, and the ontological and epistemological assumptions that underpin it. I have started this Chapter by considering data leaks as *empirical occasions* (Marres 2013); that is, as moments that can be repurposed for social scientific research so that the framing of data becomes discernible and the operations of framing devices more legible for analysis. I have considered two different methods through which such repurposing could be achieved: that of the *experiment* and that of the *case,* respectively. I have firstly considered whether the leaks that the thesis repurposed could be taken as a type of S*ociety's experiment* (Morgan 2013). Whilst I have shown that data leaks could be methodologically approached as Society's experiments, following Morgan (2013) I have argued that data leaks can also be repurposed with the aid of other epistemic genres such as the *case*, which Morgan (2012, 2013) considers as being less reductionist than retrofitted experimental designs. I hence have suggested that one way of empirically taking advantage of data leaks in a less reductionist way is to redefine them as *cases* instead; a method which aims at opening up and extending

data leaks' affordances and generative effects and capacities. Rather than understanding cases as empirically given however, I have followed Ragin's (1992) take on the activity of *casing* instead. This has entailed thinking the empirical and the conceptual as mutually interdependent, and reflecting on the ways in which materials, concepts and research results are brought together and delimited. Casing entails creating a bounded, whole object of analysis and committing to a considerable depth of engagement with the empirical materials used to craft such a whole. I have claimed that the creation of case studies constitutes a research mode geared towards the discovery of phenomena, and that as such it can be considered as revelatory. In short, the Chapter has described how the aim of using casing as a research strategy has been to use the combination of empirical materials and conceptual ideas to advance meaningful and novel descriptions of the framing of data, and the devices that establish such framings.

Having established a theoretical framework and a research design, the thesis has then presented the two case studies that were crafted as part of the research process. The first case study – entitled *The Making and Unmaking of Personal Data* – has looked at the different devices, techniques and methods deployed by a variety of actors that frame data units as personal or impersonal. This case study was empirically grounded in the release of a database containing user's search keywords from AOL's web-based search engine, and the data framing devices and techniques which this particular leak made visible. The case study has looked at relatively new ways of configuring the relation between persons and data afforded by digital technologies like search engines, whilst also paying attention to more well-established devices like

informed consent and *anonymisation* that have historically worked to reframe data as impersonal instead.

In the first Chapter that comprises this case study, I have suggested that the personalisation of search by web search engines can be understood as a new technical mode of articulating particular versions of persons. Through personalisation techniques, search keywords become configured as an expression of subjective informational intentionality. Search, understood as an act of retrieving information, has therefore become a mode of making up persons, and has thus unleashed new possible modalities of experiencing personhood(Hacking 1986). In this Chapter, I have also argued that although the configuration of search keywords as parts of and as expressions of persons is being accomplished by the technical operations afforded by search engines, the conflation between persons and search keywords is also being negotiated in the execution of social and public experiments with data. The Chapter has firstly analysed a *public reidentification demonstration* with data as a framing technique. I have claimed that apart from composing search keywords as an expression of interiority and personal dimensionality, this technique for invoking persons out of anonymised databases could also be conceived of as a framing that naturalises and adapts techno-science into public and social life, expanding what is conceived as personal and private in society at large. I have also suggested that reidentification demonstrations operate by way of establishing *merographic connections* (Strathern 1992) between search keywords and persons, and that they can also be understood as a mechanism for the *technical redoing* of data (Goffman 1974). In this Chapter, I have also analytically engaged with a data artwork entitled *I Love Alaska* as a way of thinking about data frames, in particular the frame constituted by

the reordering of search keywords into biographical sequences. I have thought of this frame as one that renders a particular nameless biographical account possible, and as one that establishes a peculiar relation between search keywords and persons. I have claimed that this framing device enables the narration of biographical accounts without having to resort to the use of a personal name. I have also suggested that framing devices like the one engineered by the artwork *I Love Alaska* force us to think about whether the function of personal names might be in demise socially, as the artwork is evidence of the generative subjectivising and individualising force that materialises despite the deletion of personal names.

Whereas Chapter Three has looked at relatively new technical means for articulating the relations between data and persons, in Chapter Four I have analysed the work of two other, more well-established devices that have historically operated to make data impersonal instead. I have looked at the devices of *informed consent* and *anonymisation* as ways of enabling data to become framed as an entity distinct and disentangled from persons. These framings, I have suggested, facilitate the economisation of data, that is, the re-framing of data so that it becomes economically valuable. I have started this Chapter by suggesting that rather than understanding research subjects as already endowed with autonomy and decision-making capacities, we should think of informed consent as a device which frames research subjects in such a way as to make the act of giving consent into an explicitly human form of agency and decision-making. I have suggested too that, as in market exchange settings, research transactions become framed as consensual by the device of informed consent, and that, in this respect, we can also think of informed consent as a 'surrogate property contract' (Waldby and Mitchell 2006) that eventually contributes

to frame data as economically valuable. In Chapter Four I have also considered anonymisation as an economisation device. I have suggested that as it erases previous 'personal' and 'private' attachments, anonymisation aids in the process of making data more thing-like, and thus more amenable to ownership. I have further argued that anonymisation suppresses the 'author function' (Foucault 1970) and in so doing disables a particular way of being a person in relation to an idea or work, in this case as the authors of the data being extracted out of them. I have suggested that the disabling of the author function via anonymisation procedures prevents persons from owning data framed (by other framing operations) as about them.

Whereas in the first case study the focus has been on the making and unmaking of personal data and the devices and techniques which framed data as personal or impersonal, the second case study has looked at a more well-established data type instead: that of price. This case study – called *Price Formations and Statistical Compositions* – has focused on the devices and techniques at work in framing and therefore forming price as economic data, paying particular attention to how price framings delimit or expand this data type's functions. This case was empirically based on the disclosure request made by Argentinean government officials to release databases containing the details of the products, services and settings used to calculate inflation in the country. As in the first case study, I have looked at new and old framing devices that serve to configure price as economic data. The case study, which starts in Chapter Five, has taken into account studies in economic sociology and social studies of finance that analyse the material and semiotic configurations that enable market agents to become aware of and attribute meaning to the movement of prices in financial and retail market settings. Based on these insights, and also on the

articulation of an incipient conceptual relation between Pierce's theory of the sign and Goffman's account of frames, I have gone on to look at the framing techniques and devices that shape price as *pure* in the measurement and legal containment of the Argentinean official inflation indicator.

The first technique I have analysed is the product specification and identification procedures undertaken as part of the offline collection of prices for the production of the Consumer Price Index (CPI). I have shown how these procedures take advantage of the framings effected by retail outlets in order to produce a further reframing of price. My aim has been to account for the different frames operating at the same time and in the same space, which configure the aboutness of price in different ways. The second device I have analysed as part of this Chapter was that of the *Statistical Secret Act*. I have argued that the operation of this particular device obscures the different frames at work that frame retail prices. I have also suggested that the visibility of these two frames would have potentially permitted a more extensive *contexting* of price, revealing other factors that are relevant in the composition of price, such as for example socio- and geo-demographic data. Moreover, I have also suggested that secrets prevent a mode of contextualisation that would have promoted the visibility of retail prices as if the multiple frames that configure them could be seen from all positions at once (Lury 1998).

In Chapter Six, the second Chapter that comprises the case study on prices, I have looked at techniques for the digital framing of prices. I have suggested that the digital measurement of inflation frames the variation of prices in time differently to offline measures. In so doing, it changes how economic fluctuations are observed by

different agents and therefore how economic action becomes organised and coordinated. This Chapter has been based on the emergence of alternative inflation indicators as a direct or indirect response to the controversy generated over the disaggregation of CPI data in Argentina. I have firstly analysed the technique of scraping online prices carried out by the Billion Price Project. I have highlighted some of the differences in how context and price aboutness are accomplished in offline and online inflation measuring practices, arguing that scraping contributes to expand and extend the *contexting* of price in a much more undetermined way for commercial ends. I have also suggested that the composition and publication of inflation figures as singular numbers has the particular social function of coordinating the observations of agents who are invested in how the economy fluctuates. The publication of official inflation figures as composites, therefore, can be thought of as a mechanism for governing price emergencies. However, I have also shown how the commercialisation of price variation establishes a different frame for the observation of price aboutness by particular groups or actors. I have argued that such framing permits a different type of economic observation through which price emergencies may be exploited rather than governed. I have demonstrated this point by describing an alternative inflation indicator recently commercialised by the data start-up company Premise. Premise undertakes a form of data disaggregation through crowd-sourced images for the identification of price emergencies, amongst other things. Premise, the Chapter has argued, contextualises the detection of inflation emergencies as a local knowledge problem, framing data through the use of digital images in order to detect rather than govern price emergencies. I have argued that the economic value of the digital images and data sold by Premise lies in the partial and advantageous

view of the overlapping lamination of frames that defines the aboutness of retail price.

To conclude the thesis, I would like to highlight how my research has both explicitly and implicitly compared newly established ways of framing data with more traditional ones. The thesis can thus also be read as offering some indication of the changing functions of traditional social coordinates like *personal names* and *official economic indicators,* in the light of the emergence of new data types, data framing devices and analytical techniques. I would also like to conclude by offering some indication of how my research could be followed up in the future.

The first case study proposed that web-searching is becoming an important mode of making up and equipping persons, and that this new technical organisation of data also unleashes new possible modalities of experiencing personhood (Hacking 1986). However, the thesis has also shown that search keywords are emerging as a personal data type through the execution of public and social experiments with data. I have suggested that in these technical re-articulations of data, what counts as personal and private is becoming redefined. I have also suggested that through these newly emerging means of configuring what is deemed to be personal or private, the traditional function of a data type such as *personal name* is being challenged. The first case study showed that - due to data anonymisation procedures, the use of database's unique identification numbers and the emergence of data sequencing framings - the function of personal names, understood as data units that have historically configured and coordinated individual and social identities in space and time, enabled singular identification, and established particular property relations,

might be socially in demise. I have demonstrated how the organisation of a biographical narrative using other data types, like search keywords, can still signal the existence of a person. Search keywords can potentially be used as a means to identify a person without having to resort to the use of a personal name, understood as a singular denominator. Identity and singularity emerged through organising, displaying and relating multiple data types. I have also shown how the personal data economy is shaped when the function of personal names as organisers of property rights between authors and works becomes deleted via anonymisation procedures. I have suggested that what enables the conversion of data into a thing and into the potential object of property claims is not the presence of a personal name – as in the author function – but in fact its removal. In studying the framing of data the thesis has also described how the function of personal names might be changing due to the proliferation of new data types and framings.

The thesis has also suggested that the ways in which official economic indicators are released and made public can come to organise a particular mode of observing and perceiving the state of a given economy and its fluctuations. Aggregation, the thesis claimed, is one framing that enables a particular mode of collectively observing the average fluctuation of prices of a given economy. Relevant here is that through official indicators, agents can observe the state of the economy synchronically, as indicators like CPI are revealed to all agents at exactly the same time. This punctuated temporal release also enables the economy to be observed symmetrically, as it distributes information equally among agents. These techniques for framing the publication of official economic indicators allow them to function as indexes of what every agent in the economy knows to be known by others. I have suggested, contrary

to a Hayekian understanding of aggregated and centralised statistics, that these numbers thus organise and coordinate a particular way of collectively observing the economy. I have claimed that the publication and framing of official economic aggregated data serves to modulate the inferences that agents make about each others' observations of the economy. Because they are established as a reference visible and available to all agents, official economic statistics can therefore be thought of as mechanisms for the governance of price emergencies. I have suggested in the thesis that the function of official economic indicators might be changing however due to the introduction of framing devices that challenge the perceptible registers that the techniques for releasing official economic indicators have historically secured. Following Amoore (2011), I have argued that the commercialisation of official economic indicators based on *data derivatives* permits a different type of observation of the economy, one geared towards exploiting rather than governing price emergencies. I have suggested that the economic value of commercial inflation data resides in the possibility of observing the multiple frames that configure the aboutness of retail price.

Personal names and official economic indicators both function as coordinates. Personal names coordinate the identification of individuals and their identities across different spatiotemporal schemas. They also bring together and organise biographical information as pertaining to a singular individual. Official indicators on the other hand are published and communicated so that they can coordinate the observation of economic phenomena synchronically and symmetrically across a population of observers. They act as references of what everyone in the economy observes and knows. The thesis however has shown that these coordinative functions of personal

names on the one hand and official economic indicators on the other are being challenged by new data types, recording, and analytical techniques and through the establishment of new and different devices and techniques for framing data. The thesis has drawn attention to what happens when social reference points like personal names and official economic indicators are confronted by other modes of indicating personhood and inflation. It has argued that different data types and their framings have important and specific implications for the organisation of social life (Lury 2012). The thesis therefore has opened up - and partially answered - the question of what is socially implicated when long-standing common references that have historically served to synchronise social action and knowledge are displaced by other data types which do not necessarily provide a similar function - at least not as collective coordinates and references.

I would like to suggest that some other empirical occasions that have taken place while writing this thesis might prove useful resources to investigate these questions further in the future. The first empirical occasion is also a result of the crisis around the validity of official statistics in Argentina. As suggested in the thesis, in recent times economists have taken advantage of natural experiments to account for the effects of public information on the formation of economic expectations. Carrillo and Emran (2009), for example, study how public economic numbers affect the formation of inflation expectations amongst households in the country, by focusing on a programming error mistake that led to the National Bureau of Statistics in Ecuador publicising an invalid CPI figure. Similarly, some scholars claim that 'the record of Argentina's economic history is so unusual in its turbulence and conflicts and promises unfulfilled that it offers an extraordinary and unparallel laboratory' (della

Paollera and Taylor 2003, 5) for different types of social and economic investigations. Indeed, it could be argued that since the ongoing controversy over the validity of the CPI and other public indicators in Argentina, the country has lost the common frame of reference previously used to collectively observe economic, and also social (as poverty statistics are also in question), phenomena. Because of the peculiarity offered by a setting deprived of social coordinates, the Argentinean case presents itself as an empirical occasion ready to be repurposed for the social study of the function of data and numbers.

An interesting aspect of considering Argentina as a setting socially deprived of common economic coordinates is that that the function of traditional indicators such as CPI has not been completely eliminated, but has been partially replaced by novel and problematic ways of organising the collective framed observation of price variation. The programme Precios Cuidados (Cherished Prices) was rolled out in Argentina in 2013, after the CPI controversy. It is a price control mechanism that was implemented by the government to apply to a range of products, also defined as 'prices and products of reference'. The government set up price agreements with a range of producers, and the programme Precios Cuidados enrols lay citizens and consumers in the verification of such price agreements throughout retail outlets in the country. These citizen consumers are provided with a free phone number, a web page and an android application, by which they can expose producers and retailers who might be breaching the agreement. Retailers are required in stores to clearly display the products which form part of the agreement. If either producers or retailers are found to be breaching the agreement by not supplying the products agreed, not displaying the right information, or charging more than what was agreed, the

government can fine them. In a similar way to CPI, the programme contributes to collectively stabilise the experience of inflation. In making the reference prices and their eventual variation publicly visible in every retail outlet - but also by engaging citizen consumers in the verification of these prices - the programme flags specific prices, and in so doing brings into being a collective experience of inflation. Citizens and consumers who do not want to take part in the actual verification process are compelled to watch, via the programme signage system, a set of products and their prices which draw their attention to price stability rather than change. Precios Cuidados can also be thought of as a way of interfering with the memories of prices which are ultimately formative of perceptions and expectations of *personal rather than collective* inflation, an interference that is achieved by directing attention to the collective register of stabilised inflation that the government exercises in its agreements.

This collective modulation of the register of inflation is important in the Argentinean context for two reasons: firstly, because of the different hyperinflation crises that the country has been through which have sedimented an economic culture of affective emergencies and led to contagious behaviour in relation to price stability and variation, particularly expressed through hoarding episodes. Secondly, and as suggested above, this collective modulation of the experience of inflation is important in Argentina due to the controversy generated around the CPI since 2006. This controversy has made the proliferation of inflation experiences anchored in personal as opposed to collectively organised perceptions more visible and predominant. In contexts like the Argentinean one, where the contagion of behaviours which lead to hyperinflation crises is always latent, the regulation of personal inflation experiences

and their confluence with collective experiences of stable price variation are crucial. On the other hand, the programme Precios Cuidados makes us reflect on how the situational regulation of inflationary experiences can be more effective than the sole publication of official inflation indicators, in particular for the formation of rational inflation expectations in countries with a latent and socially problematic hyperinflation memory.

The second empirical occasion which is worth considering in terms of the changing function of data coordinates, in this case personal names, is the *Right to be Forgotten* ruling (c131-12) by the Court of Justice of the European Union on the 13th of May 2014. This ruling was based on a case lodged by a Spanish citizen against Google Spain and Google Inc. The citizen complained that, as part of its search results, Google made an auction notice of his repossessed home available. However, the legal proceedings concerning him and his house had been already fully resolved offline, meaning that the online reference displayed by Google was irrelevant and misleading. The citizen requested that the online data relating to him was removed from the public online search result and also deleted from Google's servers. Based on the European Data Protection Directive 1995 principles, the EU Court's *Right to be Forgotten* ruling granted citizens and consumers the right to have their personal data removed from search engines when inaccurate, inadequate, irrelevant or excessive information was publicly displayed. On a factsheet distributed by the European Commission to clarify the applicability of the ruling, it was stated that search engines would have to delete the information as soon as a request from a person affected was received. The factsheet stated that 'this would mean that a citizen, whose personal data appears in search results linking to other WebPages when a search is done with

that *person's name*, requests the removal of such links' and also importantly that 'for example, John Smith will be allowed to request Google to delete all search links to webpages containing his data, when one enters the search query "John Smith" in the Google Search Box.'[116]

It is debatable whether rulings such as *The Right to be Forgotten,* and the changes to the EU's Data Protection Directive 1995 that will be introduced later this year (2015), confirm or challenge the traditional function of a data coordinate such as a personal name. What these legal cases and rulings however attest to is that legislation is a rich empirical terrain on which to explore the relations that personal names can or cannot establish with other data and entities in the world, and the prevalence of personal names as organisers of identity and individuality in society. It appears, as the thesis has suggested, that by validating certain functions of personal names via legal regulations some rights have been granted whilst others have been abolished. The right to privacy and the right to manage our identities and selves in public and everyday life (Goffman 1956) with a certain degree of freedom and autonomy have been prevalent assumptions in the design of data legislation and so-called privacy enhancing technologies. The upholding of these rights has entailed however that our right to copyright data has been overturned, even if it is considered to be the outcome of personal creative work. New datascapes, and the different legislative acts that attempt to regulate them, present themselves as empirical occasions that make it possible to trace the different ways in which the present and future function of personal names in society is being shaped.

[116] http://ec.europa.eu/justice/dataprotection/files/factsheets/factsheet_data_protection_en.pdf, last accessed 9th September 2015.

Appendix

1.0AOL'S LEAK READ ME FILE CONTENT

```
500k User Session Collection
---------------------------------------------
This collection is distributed for NON-COMMERCIAL RESEARCH USE ONLY.
Any application of this collection for commercial purposes is
STRICTLY PROHIBITED.

Brief description:

This collection consists of ~20M web queries collected from ~650k
users over three months.

The data is sorted by anonymous user ID and sequentially arranged.

The goal of this collection is to provide real query log data that
is based on real users. It could be used for personalization, query
reformulation or other types of search research.

The data set includes {AnonID, Query, QueryTime, ItemRank,
ClickURL}.
        AnonID - an anonymous user ID number.
        Query  - the query issued by the user, case shifted with
                 most punctuation removed.
        QueryTime - the time at which the query was submitted for
search.
        ItemRank  - if the user clicked on a search result, the rank
of the
                    item on which they clicked is listed.
        ClickURL  - if the user clicked on a search result, the
domain portion of
                    the URL in the clicked result is listed.

Each line in the data represents one of two types of events:
        1. A query that was NOT followed by the user clicking on a
result item.
        2. A click through on an item in the result list returned
from a query.

In the first case (query only) there is data in only the first three
columns/fields -- namely AnonID, Query, and QueryTime (see above).
In the second case (click through), there is data in all five
columns.  For click through events, the query that preceded the
click through is included.  Note that if a user clicked on more than
one result in the list returned from a single query, there will be
TWO lines in the data to represent the two events.  Also note that
```

if the user requested the next "page" or results for some query, this appears as a subsequent identical query with a later time stamp.

CAVEAT EMPTOR -- SEXUALLY EXPLICIT DATA! Please be aware that these queries are not filtered to remove any content. Pornography is prevalent on the Web and unfiltered search engine logs contain queries by users who are looking for pornographic material. There are queries in this collection that use SEXUALLY EXPLICIT LANGUAGE. This collection of data is intended for use by mature adults who are not easily offended by the use of pornographic search terms. If you are offended by sexually explicit language you should not read through this data. Also be aware that in some states it may be illegal to expose a minor to this data. Please understand that the data represents REAL WORLD USERS, un-edited and randomly sampled, and that AOL is not the author of this data.

Basic Collection Statistics
Dates:
  01 March, 2006 - 31 May, 2006

Normalized queries:
  36,389,567 lines of data
  21,011,340 instances of new queries (w/ or w/o click-through)
   7,887,022 requests for "next page" of results
  19,442,629 user click-through events
  16,946,938 queries w/o user click-through
  10,154,742 unique (normalized) queries
     657,426 unique user ID's

Please reference the following publication when using this collection:

G. Pass, A. Chowdhury, C. Torgeson, "A Picture of Search" The First International Conference on Scalable Information Systems, Hong Kong, June,
2006.

Extracted from http://gregsadetsky.com/aol-data/, last accessed 2nd January 2013.

1.1 THELMA ARNOLD

Extracted from Barbaro and Zeller (2006, August 9), A Face Is Exposed for AOL Searcher No. 4417749. *The New York Times*. Retrieved from http://www.nytimes.com, last accessed 9th August 2015.

1.2 WHAT REVEALING SEARCH DATA REVEALS

What Revealing Search Data Reveals

AOL posted, but later removed, a list of the Web search inquiries of 658,000 unnamed users on a new Web site for academic researchers. An interview with one of those unnamed users, Thelma Arnold, combined with her data reveal what she was searching for, why and on which Web sites.

A sample of Thelma Arnold's search data released by AOL

4417749	swing sets	2006-04-24	15:39:30	4	http://www.byoswingset.com
4417749	swing sets	2006-04-24	15:39:30	9	http://www.buychoice.com
4417749	swing sets	2006-04-24	15:39:30	10	http://www.creativeplaythings.com
4417749	swing sets	2006-04-24	15:39:30	5	http://www.childlife.com
4417749	swing sets	2006-04-24	15:39:30	6	http://www.planitplay.com
4417749	that do not shed	2006-04-28	9:05:54	2	http://www.gopetsamerica.com
4417749	dog who urinate on everything	2006-04-28	13:24:07	6	http://www.dogdaysusa.com
4417749	walmart	2006-04-28	14:07:32	1	http://www.walmart.com
4417749	womens underwear	2006-04-28	14:12:28	10	http://www.bizrate.com
4417749	jcpenny	2006-04-28	14:16:05		
4417749	jcpenny	2006-04-28	14:16:49	1	http://www.jcpenney.com
4417749	tortus and turtles	2006-04-29	13:12:47		
4417749	manchester terrier	2006-05-02	9:05:31	1	http://www.manchesterterrier.com
4417749	delta	2006-05-02	11:49:26		
4417749	fingers going numb	2006-05-02	17:35:47		
4417749	dances by laura	2006-05-02	17:59:32		
4417749	dances by lori	2006-05-02	17:59:57		
4417749	single dances	2006-05-02	18:00:18	1	http://solosingles.com
4417749	single dances in atlanta	2006-05-02	18:01:13		
4417749	single dances in atlanta	2006-05-02	18:01:50		
4417749	dry mouth	2006-05-06	16:49:14	2	http://www.mayoclinic.com
4417749	dry mouth	2006-05-06	16:49:14	8	http://www.wrongdiagnosis.com
4417749	tyroid	2006-05-06	16:55:34		
4417749	thyroid	2006-05-06	16:55:44		
4417749	competitive market analysis of homes in lilburn	2006-05-14	12:14:52		
4417749	competitive market analysis of homes in lilburn	2006-05-14	12:16:17		
4417749	competitive market analysis of homes in lilburn	2006-05-14	12:16:43		

Why the search

"I was thinking about my grandchildren"

"I was looking for some."

"A woman was in the [public] bathroom crying. She was going though a divorce. I thought there was a place called 'Dances by Lori,' for singles."

"I wanted to find out what my house was worth."

Extracted from Barbaro and Zeller (2006, August 9), A Face Is Exposed for AOL Searcher No. 4417749. *The New York Times*. Retrieved from http://www.nytimes.com, last accessed 9th August 2015.

1.3 AOL'S USER 711391 SAMPLE LIST OF QUERIES

711391 can not sleep with snoring husband 2006-03-01 01:24:00
711391 cannot sleep with snoring husband 2006-03-01 01:24:07 9 http://www.wjla.com
711391 cannot sleep with snoring husband 2006-03-01 01:24:07 9 http://www.wjla.com
711391 cannot sleep with snoring husband 2006-03-01 01:33:06 1 http://www.epinions.com
711391 jackie zeaman nude 2006-03-01 15:26:27
711391 jackie zeman nude 2006-03-01 15:26:38
711391 strange cosmos 2006-03-01 16:07:15 1 http://www.strangecosmos.com
711391 mansfield first assembly 2006-03-01 16:09:20 1 http://www.mansfieldfirstassembly.org
711391 mansfield first assembly 2006-03-01 16:09:20 3 http://netministries.org
711391 reverend harry myers 2006-03-01 16:10:07
711391 reverend harry myers 2006-03-01 16:10:30
711391 national enquirer 2006-03-01 17:13:14 1 http://www.nationalenquirer.com
711391 how to kill mockingbirds 2006-03-01 17:18:11
711391 how to kill mockingbirds 2006-03-01 17:18:33
711391 how to kill annoying birds in your yards 2006-03-01 17:18:58
711391 how to kill annoying birds in your yards 2006-03-01 17:19:53 2 http://www.sortprice.com
711391 how to rid your yard of noisy annoying birds 2006-03-01 17:23:08 3 http://shopping.msn.com
711391 how to rid your yard of noisy annoying birds 2006-03-01 17:23:08 10 http://www.bergen.org
711391 how to rid your yard of noisy annoying birds 2006-03-01 17:24:35 15 http://www.saferbrand.com
711391 how do i get mocking birds out of my yard 2006-03-01 17:27:17
711391 how do i get mockingbirds out of my yard 2006-03-01 17:27:36 9 http://www.asri.org
711391 how do i get mockingbirds out of my yard 2006-03-01 17:30:14
711391 how to get rid of noisy loud birds 2006-03-01 17:30:52 3 http://www.bird-x.com
711391 how to get rid of noisy loud birds 2006-03-01 17:30:52 1 http://forums2.gardenweb.com
711391 how to get rid of noisy loud birds 2006-03-01 17:30:52 10 http://www.birding.com

711391 mansfield first assembly 2006-03-01 18:31:36 3 http://netministries.org
711391 beth moore 2006-03-01 19:42:41 1 http://www.lproof.org
711391 judy baker ministries 2006-03-01 19:49:03 2 http://www.embracinggrace.com
711391 god will fulfill your hearts desires 2006-03-01 19:59:06 10 http://www.pureintimacy.org
711391 online friendships can be very special 2006-03-01 23:09:37
711391 online friendships can be very special 2006-03-01 23:09:57
711391 online friendships 2006-03-01 23:10:24
711391 cypress fairbanks isd 2006-03-02 07:56:53 1 http://www.cfisd.net
711391 people are not always how they seem over the internet 2006-03-02 08:31:51
711391 friends online can be different in person 2006-03-02 08:32:42
711391 friends online can be different in person 2006-03-02 08:33:04 13 http://www.salon.com
711391 boston butts 2006-03-02 09:47:36
711391 community christian church houston tx 2006-03-02 16:07:53
711391 gay churches in houston tx 2006-03-02 16:08:23
711391 community gospel church in houston tx 2006-03-02 16:08:45 2 http://www.communitygospel.org
711391 houston tx is one hot place 2006-03-02 18:04:44
711391 houston tx is one hot place to live 2006-03-02 18:04:55 9 http://travel.yahoo.com
711391 houston tx is one hot place to live 2006-03-02 18:16:05 1 http://www.houston-texas-online.com
711391 texas hill country and sights around san antonio tx 2006-03-02 18:19:00 5 http://www.answers.com
711391 can liver problems cause you to loose your hair 2006-03-02 18:27:04
711391 can liver problems cause you to loose your hair 2006-03-02 18:27:30 1 http://www.askdoctrish.com

Extracted from http://rhizome.org/editorial/2009/aug/19/i-love-alaska-2008-lernert-engelberts-amp-sander-p/, last accessed 8th August 2015.

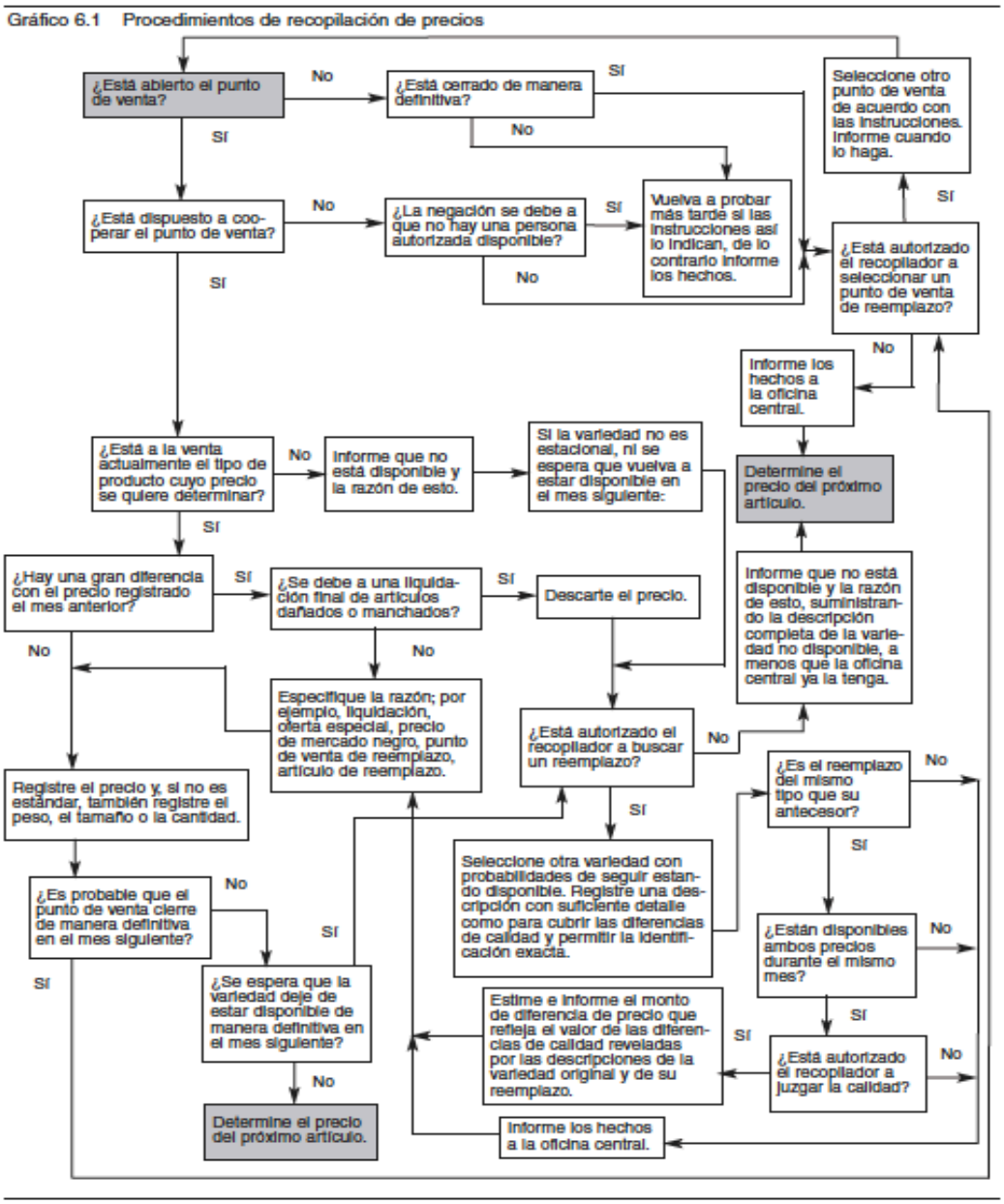
1.4 USER 1447039'S DATA EXTRACTED FROM AOL'S LEAKED DATABASE

```
14467039    mike bailey trenton nj 08648  2006-04-02 17:40:59

14467039    email directory   2006-04-02 17:41:41     7
      http://www.whowhere.com
14467039    email directory   2006-04-02 17:41:41     8
      http://www.addresses.com
14467039    email directory   2006-04-02 17:41:41     3
      http://www.theultimates.com
14467039    email directory   2006-04-02 17:41:41     5
      http://www.emailaddresses.com
14467039    etrade      2006-04-05 18:48:16     1
      http://www.etrade.com
14467039    tsp account access      2006-04-05 18:53:32     1
      http://www.tsp.gov
```

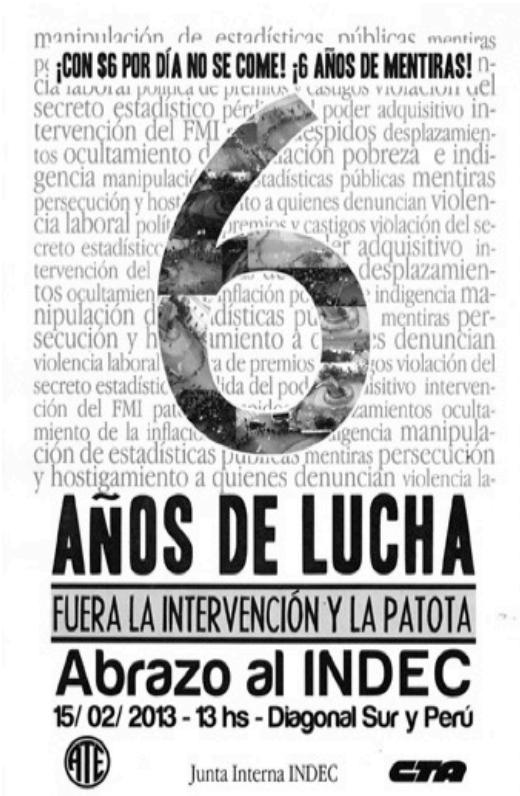
Extracted from http://gregsadetsky.com/aol-data/, last accessed 2nd January 2013.

2.0 OIT'S PROCESS OF PRODUCT SPECIFICATION



OIT Gráfico 6.1: Procedimiento Para el Anotado de Precios. *Organización Internacional del Trabajo*, IPC Manual 2006, 103.

2.1 PHOTOGRPAHIC REGISTRY OF INDEC'S DEMONSTRATIONS



Picture 1.1

Picture 1.2

Picture 1.3

Some examples of the public demonstrations that followed the controversy originated in relation to the statistical releases of INDEC in 2006*: Picture 1.1* – A poster produced by the Union, ATE/CTA, inviting members of the public to participate in an Embrace of the INDEC building. The poster reads: 'It Is Not Possible to Eat with 6 Pesos a Day! 6 Years of Lies! 6 Years Fighting! Out the Intervention and the Gang.' *Picture 1.2* – A photograph of ATE's first protest on April 10th, 2007. The banner reads, 'No to Retouched Indexes'. The capital K in 'retoKados' (retouched) refers to the first letter of the then President's surname 'Kirchner'. *Picture 1.3* – A protester at INDEC: 'Mr. President when I go shopping my wallet screams: INDEC's figures are a lie'.

All pictures are here reproduced with the permission of Claudio Rafael Castro.

2.2 DOCUMENTS TAKEN OUT OF INDEC'S PREMISES

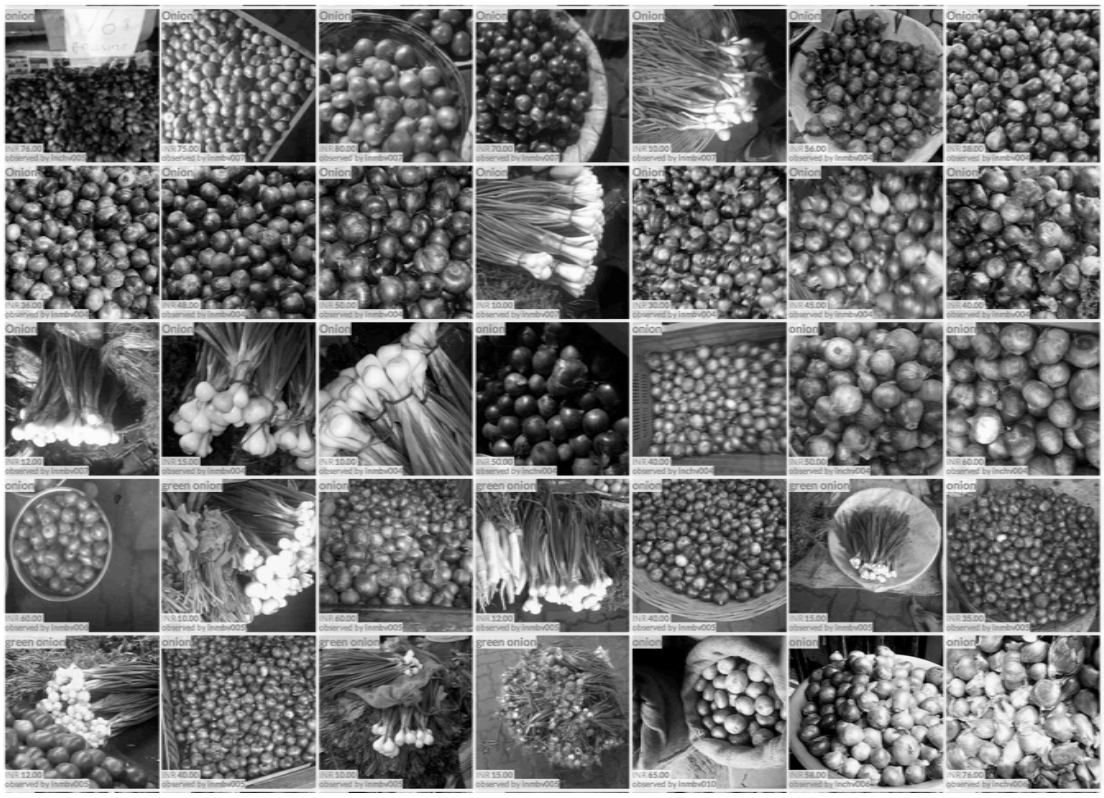
Picture 2. 1

Picture 2.2

Picture 2.1 and 2.2 – Materials are taken out of the INDEC for legal investigation, 25th July 2007.

All pictures are here reproduced with the permission of Claudio Rafael Castro.

3.0 IMAGING INDIA'S ONIONS



A crowd sourced collection of onion's images taken throughout India used to indicate price and stock emergencies.

Extracted from http://allthingsd.com/20131014/premise-creates-a-global-economic-activity-tracker-one-snapshot-of-onions-at-a-time/, last accessed 19th September 2015.

References

Abbott A. (1992). What Do Cases Do? Some Notes on Activity in Sociological Analysis. In Ragin, C. C. & Becker, H. S. (Eds.). *What Is a Case? Exploring the Foundations of Social Inquiry.* Cambridge: Cambridge University Press, 53-83.

Acquisti, A. (2010). The Economics of Personal Data and the Economics of Privacy. In OECD: *The Economics of Personal Data and Privacy: 30 Years after the OECD Privacy Guidelines*, Paris, December 2010.

Acquisti, A., Gritzalis, S., Lambrinoudakis, C. & di Vimercati, S. (Eds.) (2008). *Digital Privacy: Theory, Technologies and Practices*. New York: Taylor & Francis Group.

Adkins, L. (2005). The New Economy, Property and Personhood. *Theory, Culture & Society*, 22(1), 111–130.

Adkins, L., & Lury, C. (2009). Introduction: What is the Empirical? *European Journal of Social Theory*, 12(1), 5–20.

Adkins, L. & Lury, C. (2012). Introduction: Special Measures. *The Sociological Review*, 59(s2), 5–23.

Amoore, L. (2011). Data Derivatives: On the Emergence of a Security Risk Calculus for Our Times. *Theory, Culture & Society*, 28(6), 24–43.

Appadurai, A. (1988). *The Social Life of Things: Commodities in Cultural Perspective*. Cambridge: Cambridge University Press.

Árnason, G., & Árnason, V. (2004). Informed Democratic Consent? The Case of the Icelandic Database. *Trames,* 8(58/53), 164-178.

Asdal, K. & Moser, I. (2012). Experiments in Context and Contexting. *Science, Technology & Human Values*, 37(4), 291–306.

Back, L. & Puwar, N. (2013). A Manifesto for Live Methods: Provocations and Capacities. *The Sociological Review*, 60(S1), 6–17.

Barron, A. (2008). Copyright Infringement, 'Free Riding' and the Lifeworld. *LSE Law, Society and Economy Working Papers*, No. 17.

Bateson, G. (1955). A Theory of Play and Fantasy. Psychiatric Research Reports 2, *American Psychiatric Association*, December 1995, 39-51.

Bateson, G. (1972). *Steps to an Ecology of Mind*. Chicago: University of Chicago Press.

Bauman, R. & Briggs C. L. (1990). Poetics and Performance as Critical Perspectives on Language and Social Life. *Annual Review of Anthropology*. Vol. 19, 59-88.

Beauchamp, T. & Childress, J. (2001). *Principles of Biomedical Ethics*. Oxford: Oxford University.

Becker, V. (2013). The 2010 Census and the "Militant" Statistic. *Radical Statistics* Issue 108: 56-57.

Becker, K. & Stalder, F. (Eds.) (2009). *Deep Search: The Politics of Search Beyond Google*. New Jersey: Transaction Publishers.

Beckert, J. (2011). Where Do Prices Come From? Sociological Approaches to Price Formation. *Socio-Economic Review*, 9(4), 757–786.

Beer, D. & Burrows, R. (2013). Popular Culture, Digital Archives and the New Social Life of Data. *Theory, Culture & Society*, 30(4): 47–71.

Bell, G. (2015). The Secret Life of Big Data. In Boellstorff, T. & Mauer, B. (2015).

Data, Now Bigger and Better!. Chicago: Prickly Paradigm Press, 10-26.

Berker, T., Hartmann, M. & Punie, Y., (2005). *Domestication of Media and Technology*. Berkshire: Open University Press.

Berlant, L. G. (2000). *Intimacy*. Chicago: University of Chicago Press.

Berlant, L. G. (2007). On the Case. *Critical Inquiry*, 33(4), 663-672.

Berumen E. & Becker V. A. (2011). Recent Developments in Price and Related Statistics in Argentina. *Statistical Journal of the International Association of Official Statistics,* 27: 7-11.

Beunza, D. & Stark, D. (2012). From Dissonance to Resonance: Cognitive Interdependence in Quantitative Finance. *Economy and Society*, 41(3), 383–417.

Beunza, D., Hardie, I. & MacKenzie, D. (2006). A Price is a Social Thing: Towards a Material Sociology of Arbitrage. *Organization Studies*, 27(5), 721–745.

Boellstorff, T. & Mauer, B. (2015). *Data, Now Bigger and Better!*. Chicago: Prickly Paradigm Press.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E. & Fowler, J. H. (2012). A 61-Million-Person Experiment in Social Influence and Political Mobilization. *Nature*, 489(7415), 295–298.

Bourdieu P. (1986). The Biographical Illusion. In du Gay P., Evans J. and Redman P. (Eds.), *Identity: A Reader*. London: Sage, 297-303.

Bowker, G.C. (2005). *Memory Practices in the Sciences*. Massachusetts: MIT Press.

Bowker, G.C. & Star, S.L. (1999). *Sorting Things Out: Classification and its Consequences*. Massachusetts: MIT Press.

Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679.

Braun, B. (2015). Governing the Future: the European Central Bank's Expectation Management during the Great Moderation. *Economy and Society*, 1–25. http://doi.org/10.1080/03085147.2015.1049447

Bronk, R. (2013). Hayek on the Wisdom of Prices: A Reassessment. *Erasmus Journal for Philosophy and Economics*, 6(1), 1–26.

Brown S. D. (2012). Experiment: Abstract Experimentalism. In Lury, C. & Wakeford, N. (2012). *Inventive Methods: The Happening of the Social*. London: Routledge, 62-75.

Bruno I., Didier E. & Vitale T. (2014). Statactivism: Forms of Action Between Disclosure and Affirmation. *Partecipazione e Conflitto,* 7(2), 198-220.

Byrne, D. & Ragin, C. C. (2009). *The SAGE Handbook of Case-based Methods*. London: Sage.

Calişkan, K., (2010). The Meaning of Price in World Markets. *Journal of Cultural Economy*, 2(3), 239–268.

Çalışkan, K., & Callon, M. (2009). Economization, Part 1: Shifting Attention from the Economy towards Processes of Economization. *Economy and Society*, 38(3), 369–398.

Çalışkan, K. & Callon, M. (2010). Economization, Part 2: A Research Programme for the study of markets. *Economy and Society*, 39(1), 1–32.

Callon, M. (1986). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St. Brieuc Bay. Power. in Law J. (Ed.), *Power, Action and Belief: A New Sociology of Knowledge?* London: Routledge, 196-223.

Callon, M. (1998). *The Laws of the Markets*. London: Wiley-Blackwell.

Callon, M. (2007). What Does it Mean to Say that Economics is Performative?. In MacKenzie, D. A., Muniesa, F., & Siu, L. (2007). *Do Economists Make Markets? On the Performativity of Economics*. Princeton: Princeton University Press, 311-361.

Callon, M. & Muniesa F. (2005). Peripheral Vision. Economic Markets as Calculative Collective Devices. *Organisation Studies*, 26(8), 1229-1250.

Callon, M., & Law J. (2005). On Qualculation, Agency, and Otherness. *Environment and Planning D: Society and Space* 23(5), 717 – 733.

Callon, M., Millo, Y. & Muniesa, F. (Eds.) (2007). *Market Devices*. London: Wiley-Blackwell.

Callon, M. & Rabeharisoa, V. (2004). Gino's Lesson on Humanity: Genetics, Mutual Entanglements and the Sociologist's Role. *Economy and Society*, *33*(1), 1–27.

Carrillo P. & Shahe Emran M. (2009). *Public Information and Household Expectations in Developing Countries: Evidence From a Natural Experiment*. Institute for International Economic Policy Working Paper, The George Washington University.

Castelle M. (2013). Relational and Non-Relational Models in the Entextualization of Bureaucracy. In *Penser l'écosystème des données: Les enjeux scientifiques et politiques des données numériques,* Institut des Sciences de la Communication du CNRS (ISCC), Paris, February 2013.

Cavallo, A., Cruces G. & Perez-Truglia R. (2014). Inflation Expectations, Learning and Supermarket Prices: Evidence from Field Experiments. *NBER Working Paper,* No. 20576.

Cavallo, A. (2010). *Scraped Data and Sticky Prices*. PhD Thesis, MIT Sloan.

Cavallo, A. (2012). Online and Official Price Indexes: Measuring Argentina's

Inflation. *Journal of Monetary Economics*, 60(2), 152- 165.

Clifford, J. (1978). Hanging Up Looking Glasses at Odd Corners: Ethnobiographical Prospects. In Aaron, D. (Ed). *Studies in Biography*. Massachusetts: Harvard University Press, 41-57.

Cochoy F. (2007a). A Sociology of Market-Things: On Tending the Garden of Choices in Mass Retailing. In Callon, M., Millo, Y. & Muniesa, F. (Eds.) (2007). *Market Devices*. London: Wiley-Blackwell, 109-130.

Cochoy, F. (2007b). A Brief Theory of the `Captation' of Publics: Understanding the Market with Little Red Riding Hood. *Theory, Culture & Society*, 24(7-8), 203–223.

Cochoy, F. (2008). Calculation, Qualculation, Calqulation: Shopping Cart Arithmetic, Equipped Cognition and the Clustered Consumer. *Marketing Theory*, 8(1), 15–44.

Cochoy, F. & Grandclement-Chaffy C. (2005). Publicizing Goldilocks' Choice at the Supermarket. The Political Work of Shopping Packs, Carts and Talk. In Latour B. & Weibel P. (Eds.). *Making Things Public: Atmospheres of Democracy*. Massachusetts: MIT Press, 646-660.

Collins, H. M. (1988). Public Experiments and Displays of Virtuosity: The Core-Set Revisited. *Social Studies of Science*, *18*(4), 725–748.

Deladalle, G. (2000)**.** *Charles S. Peirce's Philosophy of Signs: Essays in Comparative Semiotics. Indiana: Indiana University Press.*

della Paollera, G. & Taylor, D. (2011). *A New Economic History of Argentina*. Cambridge: Cambridge University Press.

DeLanda, M. (2002). *Intensive Science and Virtual Philosophy*. London: A&C Black.

Desrosières, A. (1998). *The Politics of Large Numbers: A History of Statistical Reasoning.* Massachusetts: Harvard University Press.

Desrosières, A. (2014). Statistics and Social Critique. *Partecipazione e conflitto: The Open Journal of Sociopolitical Studies*, 7 (2): 330-341.

Dewulf, A., Gray, B., Putnam, L., Lewicki, R., Aarts, N., Bouwen, R. & van Woerkum, C. (2009). Disentangling Approaches to Framing in Conflict and Negotiation Research: A Meta-Paradigmatic Perspective. *Human Relations*, 62(2), 155–193.

Didier, E. (2005). Releasing Market Statistics. In Latour B. and Weibel P. (Eds.), *Making Things Public: Atmospheres of Democracy*, Massachusetts: MIT Press, 638-642.

Dilley, R. (1999). *The Problem of Context*. London: Bergham Books.

Edwards, L. (2004). Taking the "Personal" Out of Personal Data: *Durant v FSA* and its Impact on the Legal Regulation of CCTV. *Scripted*, 1(2), 342-349.

Edwards, P.N. (2010). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Massachusetts: MIT Press.

Espeland, W. N., & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, 113(1), 1–40.

Esposito, E. (2012). The Structures of Uncertainty: Performativity and Unpredictability in Economic Operations. *Economy and Society*, 42:1, 102-129.

Esposito, E. (2013). Economic Circularities and Second-Order Observation: The Reality of Ratings, *Sociologica*, 2.

Feuz, M., Fuller, M. & Stalder, F. (2011). Personal Web Searching in the Age of Semantic Capitalism: Diagnosing the Mechanisms of Personalisation. *First Monday*, *16*(2).

Finch, J. (2008). Naming Names: Kinship, Individuality and Personal Names. *Sociology* 42(4), 709-725.

Foucault, M. (1970). What is an Author?. In A. Davidson (Ed.) (2008), *The Government of Self and Others - Michel Foucault Lectures at the Collège de France*. New York: Palgrave MacMillan.

Foucault, M. (1978). The Confession of the Flesh. In C. Gordon (Ed.) (1980), *Power/Knowledge: Selected Interviews and Other Writings*, New York: Palgrave MacMillan.

Fourcade, M. (2009). The Political Valuation of Life. *Regulation & Governance*, *3*(3), 291–297.

Franklin, S. (2003). Re-thinking Nature—Culture: Anthropology and the New Genetics. *Anthropological Theory*, *3*(1), 65–85.

Fraser, M. (2008). Facts, Ethics and Events. In C. Bruun Jensen and K. Rödje (Eds.). *Deleuzian Intersections in Science, Technology and Anthropology*. Oxford: Berghahn Books.

Fuller, M. & Goffey, A. (2012a). Digital Infrastructures and the Machinery of Topological Abstraction. *Theory, Culture & Society*, *29*(4-5), 311–333.

Fuller, M. & Goffey, A. (2012b). *Evil Media*. Massachusetts: MIT Press.

Gal, S. (2002). A Semiotics of the Public/Private Distinction. *Differences: a Journal of Feminist Cultural Studies*, *13*(1), 77–95.

Galison, P. (2004). Removing Knowledge. *Critical Inquiry*, *31*(1), 229–243.

Garfinkel, H. (1967). *Studies in Etnomethodology*. Massachusetts: Blackwell Publishers.

Gitelman, L. (Ed.) (2013). *Raw Data Is an Oxymoron*. Massachusetts: MIT Press.

Goffman, E. (1956). *The Presentation of Self in Everyday Life*. London: Penguin.

Goffman, E. (1974). *Frame Analysis: An Essay on the Organization of Experience*. Boston: Northeastern University Press.

Gross, M. & Krohn, W. (2005). Society as Experiment: Sociological Foundations for a Self-experimental Society. *History of the Human Sciences*, *18*(2), 63–86.

Guggenheim, M. (2012). Laboratizing and De-Laboratizing the World: Changing Sociological Concepts for Places of Knowledge Production. *History of the Human Sciences*, 25(1), 99–118.

Guyer J. (2010). Composites, Fictions and Risk; Toward an Ethnography of Price. In Hann C. and Hart K (Ed.), *Market and Society: The Great Transformation Today*. Cambridge: Cambridge University Press, Kindle Version.

Hacking I. (1986). Making Up People. In Heller, T. C., Sosna M. & Wellbery D. E. (1986). *Reconstructing Individualism: Autonomy, Individuality and the Self in Western Thought*. California: Stanford University Press, 222-237.

Hagberg, J. & Kjellberg, H. (2015). How Much is It? Price Representation Practices in Retail Markets. *Marketing Theory*, *15*(2), 179–199.

Hallinan, D. & Friedewald, M. (2015). Open Consent, Biobanking and Data Protection Law: Can Open Consent be 'Informed' under the Forthcoming Data Protection Regulation?. *Life Sciences, Society and Policy* 11(1), 1-36.

Halpern, O. (2012). Schizophrenic Techniques: Cybernetics, the Human Sciences, and the Double Bind. *Scholar & Feminist Online*, *10*(3).

Halpern, O. (2015). *Beautiful Data: A History of Vision and Reason since 1945.* Durham and London: Duke University Press.

Hayek, von, F. A. (1945). The Use of Knowledge in Society. *The American Economic Review*, 35(4), 519-530.

Helmreich, S. (2008). Species of Biocapital. *Science as Culture*, *17*(4), 463–478.

Heritage, J. (1984). *Garfinkel and Ethnomethodology*. London: John Wiley & Sons.

Hopwood, N. (2010). Seriality and Scientific Objects in the Nineteenth Century. *History of Science* 48(3/4), 251-285.

Hurwitz, A. (1962). Constants and Compromise in the Consumer Price Index. *Journal of the American Statistical Association*, Volume 57 (13): 813-825.

James, W. (1869). The Perception of Reality. In James W. (1950). *Principles of Psychology*, Volume 2, New York: Dover Publications.

Jaszi, P. & Woodmansee M. (2003). Beyond Authorship: Refiguring Rights in Traditional Culture and Bioknowledge. In Biagioli M. & Galison P., *Scientific Authorship: Credit and Intellectual Property in Science*, London: Routledge, 195-225.

Jueguen, F. & Bullrich, L. (2010). INDEC: *Una Destruccion con el Sello de Los Kirchner*. Buenos Aires: Edhasa.

Karpik, L. (2010). *Valuing the Unique: The Economics of Singularities*. Princeton: Princeton University Press.

Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2014). Dynamic Consent: A Patient Interface for Twenty-first Century Research Networks. *European Journal of Human Genetics*, 23(2), 141–146.

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

Knox, H., O'Doherty D., Vurdubakis T. & Westrup C. (2010). The Devil and Customer Relationship Management: Capitalism and the Performativity of the Sign. *Journal of Cultural Economy,* 3(3), 339–359.

König, R., & Rasch, M. (2014). *Society of the Query Reader: Reflections on Web Search*. Amsterdam: Institute of Network Cultures.

Kurgan, L. (2013). *Close Up at a Distance*: *Mapping, Technology and Politics*. Massachusets: MIT Press.

Latour, B. (1986). Visualization and Cognition: Thinking with Eyes and Hands. *Knowledge and Society*, 06(1).

Latour, B. (1993). *The Pasteurization of France*. Massachusetts: Harvard University Press.

Latour, B. (1999). *Pandora's Hope: Essays on the Reality of Science Studies*. Massachusetts: Harvard University Press.

Latour, B. (2007). *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.

Latour, B. (2010). Tarde's Idea of Quantification. In Candea M. (Ed.), *The Social after Gabriel Tarde: Debates and Assessments*, London: Routledge 145-163.

Law, J. (2004). *After Method: Mess in Social Science Research*. London: Psychology Press.

Law, J. & Ruppert, E. (2013). The Social Life of Methods: Devices. *Journal of Cultural Economy*, *6*(3), 229–240.

Law, J. & Urry, J. (2004). Enacting the Social. *Economy and Society*, *33*(3), 390–410.

Lezaun, J. (2006). Creating a New Object of Government: Making Genetically

Modified Organisms Traceable. *Social Studies of Science*, 36(4), 499-531.

Lezaun, J. (2006). Creating a New Object of Government: Making Genetically Modified Organisms Traceable. *Social Studies of Science, 36(4)*: 499-531.

Lezaun, J. (2007). A Market of Opinions: The Political Epistemology of Focus Groups. *The Sociological Review*, 55(s2), 130–151.

Lezaun, J., Muniesa, F. & Vikkelsø, S. (2013). Provocative Containment and the Drift of Social-Scientific Realism. *Journal of Cultural Economy*, 6(3), 278–293.

Lindemboim, J. (2013). Discourse, Legislation and Data in Argentina. *Radical Statistics* Issue 108: 54-56.

Lury, C. (1993). *Cultural Rights; Technology, Legality and Personality*. London: Routledge.

Lury, C. (1997). *Prosthetic Culture: Photography, Memory and Identity*. London: Routledge.

Lury, C. (2004). *Brands: The Logos of the Global Economy*. London: Routledge.

Lury, C. (2005). Contemplating a Self-portrait as a Pharmacist: A Trade Mark Style of Doing Art and Science. *Theory*, Culture *& Society*, 22(1), 93-110.

Lury, C. (2012). "Bringing the World into the World": The Material Semiotics of Contemporary Culture. *Distinktion: Scandinavian Journal of Social Theory*, *13*(3), 247–260.

Lury, C. & Wakeford, N. (2012). *Inventive Methods: The Happening of the Social*. London: Routledge.

Lury, C., Parisi, L. & Terranova, T. (2012). Introduction: The Becoming Topological of Culture. *Theory, Culture & Society*, *29*(4-5), 3–35.

Mackenzie, A. (2012). More Parts Than Elements: How Databases Multiply. *Environment and Planning D*, 30: 335-350.

Mackenzie, A. & McNally, R. (2013). Living Multiples: How Large-scale Scientific Data-mining Pursues Identity and Differences. *Theory, Culture & Society*, 30(4), 58–91.

Mackenzie, D. & Yuval, M. (2003). Constructing a Market, Performing Theory: The Historical Sociology of a Financial Derivatives Exchange. *American Journal of Sociology*, 109: 107-145.

MacKenzie, D., Muniesa, F. & Siu, L. (2007). *Do Economists Make Markets? On the Performativity of Economics*. Princeton: Princeton University Press.

Manning, P. (1992). *Erving Goffman and Modern Sociology*. London: John Wiley & Sons.

Manovich, L. (2001). *The Language of New Media*. Massachusets: MIT Press.

Manson, N. C. & O'Neill, O. (2007). *Rethinking Informed Consent in Bioethics*. Cambridge: Cambridge University Press.

Marres, N. (2009). Testing Powers of Engagement: Green Living Experiments, the Ontological Turn and the Undoability of Involvement. *European Journal of Social Theory*, 12(1), 117–133.

Marres, N. (2012). The Experiment in Living. In Lury C. & Wakeford N. (Eds.), *Inventive Methods: The Happening of the Social*. London: Routledge, 76-95.

Marres, N. (2013). Who is Afraid of the Green Cloud? On the Environmental Rendering of Controversy. *CSISP Working Paper*, Nr. 2.

Marres, N. & Weltevrede, E. (2013). Scraping the Social? Issues in Lived Social Research. *Journal of Cultural Economy*, 6(3), 313–335.

Maurer B. (2015). Principles of Descent and Alliance for Big Data. In Boellstorff, T. & Mauer, B. (2015). *Data, Now Bigger and Better!*. Chicago: Prickly Paradigm Press, 67-87.

McCormack, D. (2015). Governing Inflation: Price and Atmospheres of Emergency. *Theory, Culture & Society*, *32*(2), 131–154.

Merton, R. (1946). *Mass Persuasion: The Social Psychology of a War Bond*. New York: Harper.

Michael, M. (2006). *Technoscience And Everyday Life*. New York: McGraw-Hill Education.

Minsky, M. (1975). A Framework for Representing Knowledge. In Winston P.H. (Ed.), *The Psychology of Computer Vision*, New York: McGraw-Hill.

Mol, A. (1999). Ontological politics. A Word and Some Questions. *The Sociological Review*, 47(S1), 74–89.

Mol, A. (2002). *The Body Multiple*. Durham and London: Duke University Press.

Morgan, M. S. (2012). Case Studies: One Observation or Many? Justification or Discovery?. *Philosophy of Science*, *79*(5), 667–677.

Morgan, M. S. (2013). Nature's Experiments and Natural Experiments in the Social Sciences. *Philosophy of the Social Sciences*, *43*(3), 341–357.

Muniesa, F. (2007). Market Technologies and the Pragmatics of Prices. *Economy and Society*, *36*(3), 377–395.

Narayanan, A. & Shmatikov, V. (2008). Robust De-anonymization of Large Sparse

Datasets. *2008 IEEE Symposium on Security and Privacy, (SP 2008)*, 111–125.

Nissenbaum, H. & Brunton, F. (2015). *Obfuscation: A User's Guide for Privacy and Protest*. Massachusetts: MIT Press.

Ohm, P. (2009). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, Vol. 57.

Parisi, L. (2012). Speculation: A Method for the Unattainable. In Lury C. & Wakeford N. (Eds). *Inventive Methods: The Happening of the Social*. London: Routledge, 232-245.

Pasquinelli, M. (2009). Google's PageRank Algorithm: A diagram of Cognitive Capitalism and the Rentier of the Common Intellect. In Becker, K. & Stalder, F. (Eds.) (2009). *Deep Search: The Politics of Search Beyond Google*. New Jersey: Transaction Publishers, 152-163.

Peirce, C. S. (1931-58). *Collected Papers of Charles Sanders Peirce***.** Hartshorne C., Weiss P. & Burks A. W. (Eds). Cambridge: Harvard University Press.

Posner, R. A. (1981). The Economics of Privacy. *The American Economic Review 71*(2), 405–409.

Prato, M. & Stark, D. (2013). Peripheral Vision in Financial Markets: How Attention Networks Shape Valuation. *Academy of Management Proceedings*, *2013*(1), 15923–15923.

Ragin, C. C. (1992). "Casing" and The Process of Social Inquiry. In Ragin, C. C. & Becker, H. S. (Eds.). *What Is a Case? Exploring the Foundations of Social Inquiry.* Cambridge: Cambridge University Press, 217-227.

Ragin, C. C. & Becker, H. S. (Eds.) (1992). *What Is a Case? Exploring the Foundations of Social Enquiry*. Cambridge: Cambridge University Press.

Ricoeur, P. (1992). *Oneself as Another*. Chicago: University of Chicago Press.

Rogers, R. (2013). *Digital Methods*. Massachusetts: MIT Press.

Rose, N. (2006). *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century*. Princeton: Princeton University Press.

Rosenberg, N. (2013). Data Before the Fact. In Gitelman, L. (Ed.). *Raw Data Is an Oxymoron*. Massachusetts: MIT Press, 15-41.

Ruppert, E. (2009). Number regimes: From Censuses to Metrics, *CRESC Working Paper Series No. 68*, Milton Keynes: Open University.

Ruppert, E. (2012). The Governmental Topologies of Database Devices. *Theory, Culture & Society*, 29(4-5), 116–136.

Ruppert, E., Law, J. & Savage, M. (2013). Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, Culture & Society*, *30*(4), 22–46.

Savage, M. (2009). Contemporary Sociology and the Challenge of Descriptive Assemblage. *European Journal of Social Theory*, *12*(1), 155–174.

Savage, M. & Burrows, R. (2007). The Coming Crisis of Empirical Sociology. *Sociology*, *41*(5), 885–899.

Scheper-Hughes, N. (2001). Bodies for sale – Whole or in Parts. *Body and Society 7*(2–3): 1–8.

Schutz, A. (1945). On Multiple Realities. In Schutz A. (1945) *Philosophy and Phenomenological Research*, Volume 2, 533-576.

Schwarz, A. & Krohn W. (2011). Experimenting with the Concept of Experiment: Probing the Epochal Break. In Nordmann A., Radder H. & Schiemann G. (Eds.), *Science Transformed? Debating Claims of an Epochal Break*, Pittsburg: University

of Pittsburg Press.

Scoble, R. & Israel, S. (2013). *Age of Context: Mobile Sensors, Data and the Future of Privacy*. New York: Patrick Brewster Press.

Seaver, N. (2015). Bastard Algebra. In Boellstorff, T. & Mauer, B. (2015). *Data, Now Bigger and Better!*. Chicago: Prickly Paradigm Press, 27-47.

Seaver, N. (2015). The Nice Thing About Context is That Everyone Has It. *Media, Culture & Society*, 37(7): 1101–1109.

Simmel, G. (1906). The Sociology of Secrecy and of Secret Societies. *American Journal of Sociology, 11*(4), 441-498.

Smith, K. (2012). From Dividual and Individual Selves to Porous Subjects. *The Australian Journal of Anthropology*, 23(1), 50–64.

Stalder, F. & Mayer C. (2009). The Second Index: Search Engines, Personalization and Surveillance. In Becker, K. & Stalder, F. (Eds.) (2009). *Deep Search: The Politics of Search Beyond Google*. New Jersey: Transaction Publishers, 98-115.

Stark, D. (2013). Observing Finance as a Network of Observation. *Sociologica*, 2.

Stenner P. (2012). Pattern. In Lury, C. & Wakeford, N. (2012). *Inventive Methods: The Happening of the Social*. London: Routledge, 136-147.

Stigler, G. J. (1980). An Introduction to Privacy in Economics and Politics. *The Journal of Legal Studies 9*(4), 623–44.

Strathern, M. (1992). *After Nature: English Kinship in the Late Twentieth Century*. Cambridge: Cambridge University Press.

Strathern, M. (1988). *The Gender of the Gift*. California: University of California Press.

Strathern, M. (2002). Externalities in comparative guise. *Economy and Society*, *31*(2), 250–267.

Strathern, M. (2004). Losing (out on) Intellectual Resources. In A. Pottage and M. Mundy (Eds.), *Law, Anthropology and the Constitution of the Social: Making Persons and Things*, Cambridge: Cambridge University Press, 201-234.

Suchman, L. A. (1987). *Plans and Situated Actions*. Cambridge: Cambridge University Press.

Sweeney, L. (2002). k-Anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 557–570.

Sweeney, L., Abu, A. & Winn, J. (2013). Identifying Participants in the Personal Genome Project by Name. *SSRN Electronic Journal, Harvard University.*

Tackz N. (2012). *Wikipedia and The Politics of Openness*. PhD Dissertation: The University of Melbourne.

Taylor, L., Schroeder, R. & Meyer, E. (2014). Emerging Practices and Perspectives on Big Data Analysis in Economics: Bigger and Better or More of the Same?. *Big Data & Society*, *1*(2), 1–10.

Thrift, N. (2004). Movement-space: The Changing Domain of Thinking Resulting from the Development of New Kinds of Spatial Awareness. *Economy and Society*, *33*(4), 582–604.

Tober, D. M. (2001). Semen as Gift, Semen as Goods: Reproductive Workers and the Market in Altruism. *Body & Society*, *7*(2-3), 137–160.

Uprichard, E., Burrows, R. & Parker, S. (2009). Geodemographic Code and the Production of Space. *Environment and Planning A*, 41(12), 2823–2835.

Venturini, T. & Latour B. (2009). The Social Fabric: Digital Traces and Quali-Quantitative Methods. In Proceedings of Future En Seine, Paris.

Vesna V. (2007). *Database Aesthetics: Art in the Age of Information Overflow*. Minnesota: University of Minnesota Press.

Waldby, C. (2000). *The Visible Human Project*. London: Routledge.

Waldby, C. (2002). Stem Cells, Tissue Cultures and the Production of Value. *Health*, 6: 305–323.

Waldby, C. & Mitchell, R. (2006). *Tissue Economies*. London and Durham: Duke University Press.

Weizman, E. (2014). Introduction: Forensis. In Forensic Architecture (Eds.), *Forensis: The Architecture of Public Truth*. London: Stenberg Press.

Williams, W. C. (1963). *Paterson*. New York: New Directions Publishing.

Zelizer, V. A. (1978). Human Values and the Market: The Case of Life Insurance and Death in 19th Century America. *American Journal of Sociology*, *84*(3): 591–610.

Zelizer, V.A. (1985). *Pricing the Priceless Child*. Princeton: Princeton University Press.