The Coalescent State: Assemblages of Surveillance and Public Policy

CHRISTOPHER CAMPBELL

Submitted in fulfilment of the requirements of the University of Stirling for the degree of Doctor of Philosophy

February 2020

Acknowledgements

An ongoing project like this bears a debt of gratitude to a large number of people. Prof. William Webster should take the lion's share of that gratitude for his ongoing support, encouragement to come back to Stirling to undertake this research in the first place, proof-reading of journal and conference articles, and guiding me through the maze of EHRC and DTC funding. Thanks also to Kirstie Ball and Charles Raab, the remainder of the tripartite supervision arrangement we put in place for me as the first CRISP doctoral student.

I would also like to thank Dr. Scott Brown, Georgia Institute of Technology, for his ongoing insight into the process of creating (and finishing) a PhD in the face of life in general, and Rosamunde Van Brakel at Vrije Universiteit Brussel, collaborator on a number of ideas and talented co-author on my first journal article, for making me feel less like I was shouting into the void with the ideas contained within this thesis.

Of course, I must also thank my family, my wife Maria who has absolutely and resolutely put up with the, at-times, nonsensical workload I have inflicted on myself over the last five years, and who has propelled me across the line at a metaphorical 11:59pm on the final day of the final month of the final year. John, my incredible boy, who has never known an existence without Dad 'writing a book': you have been an inexhaustible source of smiles, and incredibly patient with all of that.

Part One:

Introduction and

Background

Chapter 1 : Introduction

Surveillance has become, and remains, the dominant organising principle of late capitalist modernity. The methodologies and practices of surveillance reach into every aspect of society, from the local and personal to the political and macro. In the last ten years, the surveillance systems that underpin the modern state have been repeatedly exposed, to the surprise of most citizens, and to the grim familiarity to those who study surveillance. The extent to which surveillant mechanisms underpin specific practices of surveillance – and the machinery of governments more generally – had been hinted at, but never confirmed so boldly. This research seeks to evaluate the critical and conceptual frameworks that are used to discuss policy and surveillance, and to build a new framework that explicitly focuses on the complex, reflexive nature of those discussions.

This research attempts to create an enhanced critical framework for surveillance and polity using assemblage theory as its theoretical basis. It argues that existing approaches to policy and policymaking are insufficient, particularly where they relate to surveillance and consequential issues. As heuristics and systems that rely on their own internal rationality, they are ill-equipped to provide timely and usable critical insight to modern policy problems. The research goes on to try and understand how this shapes discourse and individual responses and experiences, by examining public discussion and dialogue centring on protests and surveillance.

Secondly, this research makes use of assemblages as both a unit of analysis, and as a method of critical analysis that obviates some of the difficulties of analysing complex topics which transcend the borders between differing scales of analysis. Assemblages are multiplicities that make up a whole. They are an organismic metaphor for systems and relations. Rather than assuming the existence of a single holistic whole for any specific process, organisation or system, assemblages are systems that are characterised by

the relations of exteriority between the heterogeneous components that compose them (Dosse, 2012:138). These components of assemblages are referred to as *assemblants* in the terminology of this research.

Assemblages are emergent and intensive. Emergence indicates that the characteristics of the assemblage emerge from the performance of the interaction of the constituent assemblants, rather than from a reified understanding of the components' attributes. The notion of being intensive in a Deleuzian frame indicates that a characteristic of the assemblage cannot be altered or divided without changing the characteristics and nature of the system.

This research also argues that discourse is a key component of the acceptance and dissemination of assemblages, and that the disparate levels of operation of 'framing' should be brought together with a consistent terminology. This should regard discourse not as isolated utterances or instances but instead as an expressive part of a wider assemblage, allowing research to more effectively consider the relationship between policy and discourse, in the context of that wider assemblage.

The aim of this doctoral research is to provide critical insight into surveillance policy by identifying and mapping the discursive aspects of public policy assemblages in place around protest in the UK in 2015.

Surveillance reflects the asymmetrical application of power, whether economic, social or otherwise. It is a coercive act, and is never benign (Fuchs, 2011). Surveillance is an act of structural violence because of the systematic way it reflects capitalist power structures, and the manner in which it segregates those lesser used to the mechanic of those structures from those who are.

Across most definitions of surveillance, the asymmetry of power between surveillor and surveilled is apparent. It is a systematic emanation of the capitalist state assemblage, one which both constitutes and represents it.

From a strictly utilitarian point of view, it is interesting to assess the 'benefits' of surveillance against the costs and damage to agency, liberty and privacy that it brings. Because surveillance is intrinsically linked to mechanisms of structural control, it represents a form of structural violence – another "avoidable impairment of fundamental human needs" (Galtung, 1969). Most benign or 'positive' aspects of surveillance should instead be gathered under a definition of monitoring, rather than surveillance. Surveillance, then, is the asymmetric application of power, not just through the act of watching, but through the consequential actions of that watching. The *act of watching* in surveillance is inseparable from either the reason for watching or the impetus for decision-making and action that results from the watching.

The research objectives and questions considered by the research are:

1. What discursive assemblages are in place around surveillance (particularly at protests), and how do they persist and replicate across different contexts?
2. What are the characteristics of those discursive assemblages and how do they change over time in relation to 'real-world' events?
3. How can existing social theory be utilised to adequately analyse and provide insight into the operation of assemblages?
4. What does operationalising the assemblage in this way have to offer in terms of critical understanding of public policy and surveillance?

The focus of the study changed over the course of the research, and the original approach, and research questions, are laid out in Chapter Six. Theoretical work, and the practical experience of carrying out fieldwork which suggested that some of the original questions were poorly formed, or inconsistent with the findings from developing the theoretical approach. As a result, the research questions were reframed, and a new approach, focused on analysing discursive assemblages through social media data, was formulated and adopted. The shift to this new approach, along with a detailed explanation of the background to the topic, is laid out in Chapter Seven.

Discursive assemblages of surveillance

This research makes extensive use of the concept of assemblages, drawn from the work of Deleuze and Guattari, particularly as developed in *A Thousand Plateaus: Capitalism and Schizophrenia* (Deleuze and Guattari, 1987)

Assemblages are multiplicities which make up a whole. They are an organismic metaphor for systems and relations, where the overall nature of the system is characterised by the relationship between its heterogeneous parts (Dosse, 2012:138), and defined by relations of exteriority. If a part is removed from the system and placed in another, the relations both between the part and the new system, as well as the remaining parts of the existing system, is altered, which makes them *intensive processes.*

As outlined in Chapter Five, the relations of components within an assemblage can (initially) be considered on two main axes: the material and the territorial. The first axis reflects whether the relationships inside an assemblage *affect* something, such as behaviour or whether they are expressive and lead to what can be termed discursive emanations - fragments of discourse relating to, or resulting from, the assemblage within which they were generated. The second axis examines whether the operation of that assemblage will tend to solidify the assemblage (territorialisation[1]) or makes it less stable and more open to change (deterritorialisation).

At the time of design and fieldwork, there were comparatively few large scale analyses of social media data. This research gathers millions of data points, and offers an achievable and at least partially replicable route to undertaking data research on a relatively large scale for projects with limited resources.

1

Surveillance as a Subject of Enquiry

Surveillance is the dominant method by which late capitalist society is organised. This ranges from government bodies and police forces to companies. State level surveillance socially sorts citizens into categories defining citizenship, access to services and welfare, while companies of all sizes utilise behaviour tracking as their primary method for gathering data on users, generating data doubles of users (Haggerty, 2005) to be repackaged and sold.

As a field, surveillance studies is a comparatively new field, but is an interdisciplinary field and approach, encompassing the social and political sciences. It is, at its heart, a study into the mechanisms and processes of power, whether from a societal, political, social or technical perspective (Ball and Haggerty, 2005).

Governance, as currently arranged requires classification and delineation between the people present in society into ever-more granular categories. At a governmental level, societies are divided into citizens and non-citizens, with attendant rights, requirements and legalities flowing from the categorisation. Internally, there are a myriad, or indeed, multiplicity of overlapping and sometimes competing categorisations, giving life to data-doubles (Haggerty and Ericson, 2000), that can have as much eventual influence over an individual's activities and the availability of particular routes through society as their own actions.

The *a priori* categorisation of processes or things as surveillance is a danger identified by Ball and Haggerty, one which potentially does nothing to understand the mechanism and lived experiences of those subjected to surveillance systems.

Instead, this research sets out a theoretical framework that allows for increasingly reflexive critical review of surveillance, focused on the '*control, resistance, emergence and development of surveillance practice.*' (Ball and

Haggerty, 2005:133). Surveillance is an act of structural violence situated and sited by its praxis, where the ongoing effects and drive to surveill are continually recreated. As surveillant systems and assemblages overlap and interact, they create emergent properties that exceed the properties of individual systems and components. The focus for this research is therefore on the interactions of elements that make up potentially surveillant systems, in conjunction with their experienced values and outputs.

Aims and Objectives

The core aim of this research is to advance a new theoretical framework that significantly enhances critical theory in the domain of policy studies and surveillance studies. This is achieved through an examination of the existing tools of policy, and their shortcomings, and proceeding through those problems under this new framework. Fieldwork for the study is an investigation of large-scale discourse around protests, which have a strong relationship with surveillance and surveillant acts. In undertaking this fieldwork, the research also offers some enhanced tools and emerging methodologies for gathering and implementing analysis on a large scale.

These aims contain several objectives:

- Enhance critical frameworks for policy and policy analysis in an increasingly complex world.
- Create, manage and document how large-scale research can be operated in social science, using social media.
- To offer explanations for current patterns of discussion, awareness and intent around surveillance in public spaces and protests.

Focus of the Research

The central focus of the theoretical basis for this research is the public policy making processes, and how they relate to an increasingly complex world

dominated by surveillance which is asymmetrical, ubiquitous, and which entrenches existing inequalities and power dynamics.

Rather than focusing on the traditional steps that, to a greater or lesser degree, have been the focus of most policy analysis, this research focuses on the position of policy and complexity. This research argues that governments increasingly do not 'govern', as has been traditionally understood. Accordingly, this research examines the underlying processes by which that statement is both held to be the case, and also why it is the lack of agency within government and policy analysis has been a non-obvious statement for analysts, and has been, a non-obvious statement for policy analysis and academic studies.

This then turns to look at postmodernist theory, often overlooked or derided in policy analysis, to consider how that level of complexity – and arguably lower levels of state-centric agency – can be differently conceptualised through the use of assemblage theory.

To operationalise this approach, the research examines a large volume of public discourse, gathered from social media, focusing around two protests. London is one of the most heavily surveilled locations in the Western world, and protests are doubly so. The Metropolitan police have had a Social Media Intelligence (SOCMINT) division monitoring social media for a number of years, and this research uses some of the open source technology available to both replicate that monitoring on a smaller scale, and to examine discourse and viewpoints around protests, which represent a particular hyper-surveilled location.

Choice of Research Topic/Method – Case studies

> "*Tilly sees public demonstrations as large-scale conversations between a movement, a countermovement, and the police.*" (DeLanda, 2006:87)

Following the 2008 worldwide financial crisis, there were a number of significant protest and activist movements in the UK, as well as around the world. Many of these protests were held to be facilitated by social media and technological means - civil activism and protest being enacted through the same mechanisms that are used to surveill populations. From the Arab Spring in 2011 to Occupy Wall Street and the student protests or 2011's London riots in the UK, social media was credited and blamed in equal measure for its role in civil movements and disturbances.

From the perspective of practical research, these protests are significant in having both episodic and thematic aspects. They are episodic in that they occur at specific times, with an often well-documented build up, occur over well defined geospatial and temporal windows, and have corresponding documentation and coverage or public conversations in the media and on social media. They are thematic in that protests emerge (literally and in a deleuzian sense) from the ongoing structural reasons (in terms of policy, action and the body politic) that triggered some kind of a response from the activist movement.

The seeming upturn in protest and organised movements coincides with the growth of social media generally, and particularly with the emergence of Twitter as a global platform for debate, discussion and dissent. This growth in social media and publicly available discourse is mirrored in the increasing amount of academic interest devoted to online communities and the discourses they partake in. The final factors in steering the choice of topic were a rise in levels of awareness around surveillance and privacy at a number of levels, and networked activism emerging as a topic for research and academia in the late 2000s. At the time of research, online communities have arguably moved from being a separately regarded constituency of people towards being more reflective of wider society, through the large scale uptake of social media platforms:

> "*Media technologies in academic discourse, particularly when addressed as media practices, are not clearly identifiable machinery or*

> *tools developed from scientific knowledge, but are semantically rich concepts in which the technical aspect (the technology itself) is entangled from (sic) social expectations regarding the technology, research methodologies, and a common academic platform*"
> (Neumayer and Rossi, 2016)

Significance of the Research

The research provides a contribution in two main areas. The first is a focused expansion of the concept of assemblages as a critical tool, and particularly how relationships and interactions between organisations, individuals and their respective domains can be better understood. Rather than being reliant on domain-specific heuristics, as has traditionally been the case, assemblage theory in public policy provides a context-sensitive, scale appropriate system of analysis that can operate across domains and jurisdictions. This research advances the study of policy, specifically the subjectivity of surveillance policy around protests.

The second area is around tooling and methodology. At the time of design and fieldwork, it was relatively unusual to see large-scale analyses of public discourse, except in larger and well-funded projects. This research gathers analyses millions of points of data, in attempt to understand public discourse around surveillance and protest. Discourse is a central, constitutive element of any assemblage, and offers a well-worked example of how single researchers can analyse data at large scale, while linking those research outputs to the theoretical question at hand: *How can assemblages of discourse and public policy be used to address questions of public policy and surveillance studies*?

Thesis Plan

This thesis is divided into four main parts. Part one consists of the introduction and background to the thesis. Part two encompasses the

literature review, while part three covers the research methodology, research and analysis sections. Part four, the final section, offers concluding comments, theoretical implications of the fieldwork, and suggestions for further work.

Part Two, the literature review, is divided into four chapters. Chapter Two presents a review of the literature that seeks to understand policy and the policy process, covering the evolution of the field, and outlining where the usefulness of current models and heuristics may end. It argues that, in spite of a significant body of literature, current methods and frameworks for policy analysis are insufficient to adequately capture the complexity of the modern policy world.

To address this, Chapter Three examines Foucault's Governmentality as an approach to public policy. Although Governmentality is more reflexive and self-critical than many policy process models, its focus is too much on the genealogy of the moment to provide a significant toolset for policymakers.

Chapter Four examines definitions and approaches to understanding surveillance, particularly in terms of the complexity and interrelationships that it brings to the modern environment. Crucially, this chapter positions surveillance as an act of structural violence, by examining the intentions, purpose and materiality of surveillant actions, to differentiate surveillance from more benign forms of watching, which are instead classified as monitoring or guardianship.

Chapter Five takes these comparatively disparate threads and considers their relationship, and how they can be best conceptualised through the lens of multiplicities, or assemblages of policy. Instead of conceptualising policy as a thing, or even as a specific, repeatable process, this chapter lays out how assemblages are created and recreated constantly, and how policies may be the metastable output of that constant recreation. Constructing the framework in this way allows us to consider the 'state' not as a unitary object, but as a coalescing assemblage, which crystallises around particular decision points

and outputs, in ways that can be broadly, but not consistently, understood in terms of the assemblants in place at the time. It argues that policy is immanent to its assemblages, which are greatly under-represented in policy thinking and planning.

Part Three of the thesis presents the fieldwork. Chapter Six outlines the original approach to the research (from a theoretical perspective), outlines the challenges that were faced in carrying out the research, and reflects on how those challenges could have been better mitigated during the lifetime of the project. Chapter Seven goes on to lay out a new approach, as well as a detailed rationale for the topic and subject selection, and details how the final practical approach was implemented.

Chapter Eight explores the data that was collected, and presents the analysis of two large-scale protest events in the UK. At the time it was carried out, this research was an outlier in terms of its large scale, and particularly so in terms of single-researcher projects. Consequently, it also highlights some of the practical difficulties in large-scale analyses, and areas where such studies can be streamlined.

Part Four of the thesis evaluates the methodological and theoretical implications, with Chapter Nine offering areas for further work and development, and evaluates the contribution made in terms of meeting the research objectives. A number of concluding comments are given.

Firstly, it notes that the research has identified a potentially invaluable theoretical framework for considering policy actions and connecting them to the lived subjectivity of policy subjects: Assemblage theory can provide a reflexive and recursive framework for analysis.

Secondly, however, it notes that this framework is difficult to operationalise. Although fieldwork has highlighted the potential in how research can understand changes to the shape and velocity of online discourse, the challenges in tieing fieldwork back to the specific surveillance topic in hand

have restricted the ability to draw any wider conclusions, and instead offers extensive suggestions about how these limitations can be mitigated in future work.

Thirdly, the methodology for fieldwork shows that, with more preparation and a more nuanced approach – particularly around understanding sentiment of discourse – it is entirely feasible, and indeed, probably desirable, to operate large-scale, on-going research projects centred around publicly available discourse. Where possible, these discourse analyses should be linked back to mainstream press articles, which provide a more stable textual corpus for comparison.

Part Two:

Literature Review

Chapter 2:

The Public Policy Process

Introduction

The purpose of the next three chapters is to situate the thesis in relation to existing knowledge about governance, surveillance and epistemology. These first two chapters explore the limits of current or recent public policy thinking, what surveillance means and how the subjectivity of surveillance targets can be examined. The last chapter in this section lays a detailed consideration of assemblage theory and how it provides a more naturalistic and reflexive way of approaching public policy, in light of the preceding two chapters.

Surveillance has risen to become a dominant way of life, one where the apparatus of surveillance, its practices – and the philosophies that lead to and stem from surveillance – have become the dominant force in polity, policy and politics. Public debate and discussion has been predominantly binary, while policy discussions are generally focused on the application of – or minimising the visibility of – surveillance. Surveillance systems underpin and facilitate government at every level, and they form an intrinsic part of public policy thinking.

Public Policy and the Limits of Unitary Rationality

"Running through much of the modern work that is being done on the decision process is the desire to abolish discretion on the part of the chooser and to substitute an automatic machine-like routine." (Lasswell, 1955:387)

Policies are the rules and principles that govern how organisations operate, how decisions are made, and how interactions should take place. It is 'how things shall be done', as well as 'how things are done': both a process and a

'thing' (Ham and Hill, 1984). Policy intends to solve a problem, or problematisation of an issue, purportedly in the public interest (Nakamura, 1987), although the nature of public interest is a heavily contested issue. Policies are of significance in social science as they encompass individual relations, institutional and organizational interactions and are held to reflect and demonstrate the structure and flows of the machinery of governance. Policy analysis, therefore, is that which seeks to illuminate policy. This can take the form of analysis *for* policy, or analysis *of* policy.

However, as is argued throughout the rest of this chapter, much of the work in the policy field seeks to demonstrate - or at least find heuristics towards - some sort of unitary rationality. This sees policy and its related processes, actors and outputs as an ideal type (Schütz, 1972:244) that is fixed an invariant. The desire to see policy as a consistent totality that is repeatable, overarching and consistent has lead to deeply held assumptions about the way that the policy process operates, and even around the *potential* for it to operate, that are at best flawed, and at worst, unable to properly consider the increasingly complex world that surrounds traditional policy institutions.

Analysis for policy collects evidence and arguments, purportedly weighing them rationally to come to a 'best' solution, while analysis of policy often focuses on the content and operation of policy, particularly on financial efficiency (Hill, 2009:5).

Public policy and policy analysis are the study of the processes that generated policy, the observable emanations of the state, generally in regard to its institutions and actors. The research tradition is drawn from Lindblom and Lasswell, Taylor and Weber, Ford and Schumpeter; accordingly, it owes an enormous debt to the rationality of time and motion studies, and of classical economics. Rational processes, with significant and impartial evidence bases, are held to not only be possible but also to be desirable and a key objective of the structures in place.

In contrast, Governmentality, a term developed by Foucault (1991, 1997) is drawn from a wider, more sociological background. It focuses on the actions and exercise of power and authority. Rather than focusing on rationalism as an objective in itself, argues that it is only one aspect of governance of the state, along with mentalities of government and practical actions.

Lastly, assemblage theory, such as that posited by DeLanda (2006) and Srnicek (2007), building on the work of Deleuze and Guattari argues that governments and systems of control are networks of networks that interact, changing through their interaction, and which tend to stratify (or territorialise) over time, becoming more solid. The visible aspects of these networks are what are generally recognised and interpreted as government and governance. These systems are often only internally coherent and rational. Interaction between differing conceptions of rationality can provide unexpected actions, while overlapping and congruent systems of rationality provide a rapid solidifying of those systems.

Defining Classical Public Policy and Its Processes

Nakamura and Smallwood (1980:31) suggest that:

"A policy can be thought of as a set of instructions from policy makers to policy implementers that spell out both goals and the means for achieving those goals."

In effect, policy is how a government governs: the practicalities of finding, considering and making decisions on particular issues that affect citizens. It is the politics, institutions, processes and people that go into decision-making. This covers not just how policy 'solves' problems that are presented to it (Birkland, 2011), but how it creates the problematisation that is then solve. Crudely put, it is how governments rule the population, whether by imposing norms and managing relationships (Vickers, 1965), or by making policy interventions (Considine, 2005). It is concerned with, by definition, issues of interest of affecting the public, the general population of a governed area.

The delineation of what constitutes private or public issues warranting attention from the state has been increasingly blurred in recent years, and becomes an important definitional debate in its own right (Birkland, 2011).

Wildavsky (1979) notes that "*there can be no one definition of public policy analysis",* and Gordon, Lewis and Young (1977:12) note the distinction between analysis of policy and analysis for policy. Analysis of policy examines policy itself, outcomes and impacts, often with the intent of critiquing that policy – or the originating government body as a whole.

In contrast, analysis for public policy is more heavily rooted in the empirically based, rationalist school of thought, and centres on gathering information to more effectively state the 'problem', and evaluate potential solutions. Its goal is to increase the capacity of decision and policy makers to problematise situations. This is seen as desirable because:

"We fail more often because we solve the wrong problem than because we get the wrong solution to the right problem". (Ackoff, 1974)

The aim is to make 'better' policy, a heavily contested term. Good policy is a term that tends to be framed by the policy-making body, rather than through understanding the lived experiences of policy 'recipients' (McConnell, 2010). It is important to retain a perspective of that lived experience, as it provides a counterpoint to the abstraction of rationality, where policy effects are reduced solely to that which is quantifiable. For individuals and communities, *"policy helps define the things a community holds to be important."* (Considine, 2005:16)

The Rise of Public Policy and Public Policy Analysis

In the late 1970s, a neo-liberal resurgence saw the post-war Keynesian settlement being challenged, particularly in the United Kingdom and the United States (Hood, 1990). Critics saw the state interventionist and welfare state approaches to policy issues as inefficient, and sought to 'modernise' the

public sector. Again, the focus on modernity and rationality is held up as a desirable outcome. The New Public Management (NPM) school (Hood, 1990) intended to bring the 'efficiencies' of the private sector to the public sector, as typified by the "three Ms": markets, measurement and managers (Ferlie *et al*, 1996).

The approach was characterised by decentralisation, management by objectives and a 'consumer' orientation (OECD, 2003). In the UK, this took the form of privatisation of utilities and heavy industry during the 1980s and 1990s, building on the efficiencies pursued by the Rayner Scrutinies (National Audit Office, 1986). These were a series of reviews commissioned by the Conservative government and carried out by Derek Rayner, the former Chairman of Marks and Spencer, who offered recommendations to run the services of government in a more 'commercial' manner. This led to the breakup of some of the larger governmental departments and the creation of the Next Steps Agencies, executive agencies within government with a *"clear focus on delivering specified outputs within a framework of accountability of Ministers."* (Cabinet Office, 2006)

New Public Management and The Third Way

NPM has in turn been challenged by 'Third Way' approaches (Giddens, 1994, 1998), Public Value Theory and network governance. The Third Way (re)advocated social democracy, rejecting both top-down socialism and an unfettered free market, being brought forward by the Clinton administration and the New Democrats in the US in the mid 1990s, and by New Labour under Tony Blair in the UK (Jessop, 2005). Giddens viewed it as a 'new progressivism', underpinned by equal opportunity, personability and active communities of engaged citizens (Giddens, 1998:15) – an answer to both large-scale statism and neoliberalism. However, even by Giddens' own admission, Third Way thinking is less of a policy framework or approach as it is an approach to the principles that underpin those frameworks – more an

expression of ideology and principles than a concretised framework or approach.

Public Value Theory (Moore, 1995) sought to codify the public service ethos, and understand the normative context of how public sector organisations *should* be organised. The movement towards focusing on *value*, rather than just cost, is in contrast to NPM, and rejects simple notions of efficiency. Instead, PVT attempts to consider '*how can the organisation make itself more valuable to the public?'.* This concept of understanding value centres around asking: *value to whom, on what bases, and at what cost?* As with earlier theories of policy analysis and processes, this remains centred on the concept of a single 'best' way to do public policy, with organisations that are heavily rooted in empirical and rationalist approaches being those that set the frameworks of what constitutes value.

Network governance theory offers a more complex model, and notes the decentralised and multi-actor nature of governance arrangements (Jones, Hesterley and Borgatti*,*1997). Being focused on economic activity and as a form (or emerging *practice)* of corporate management, it largely concentrates on the meeting of external market demands for consumer goods.

The models of activity that it produces, by considering short-term and informal networks within the corporation and the social networks of employees as a means of delivering corporate intentions, are useful in wider public policy questions. However, the difficulty of analysing and understanding such shadow organisations and structures is not addressed, and these problems are exacerbated by an order of magnitude when considering larger and multiple public sector organisations dealing with sometimes intractable public policy issues. In particular, Jones *et al.* note that '*without … appropriate supporting social mechanisms … both coordination and safeguarding are likely to suffer.*' (Jones, Hesterley and Borgatti*,*1997: 938).

Across each of these approaches, there exists a rationality, a 'best in class' mentality, although to a lesser extent in Network Governance Theory. Each

subsequent movement is built on an epistemic certainty in the surety and veracity of their methods; each creates a system with its internal rules, from which spring winners and losers as appropriate.

"Traditional policy analysis compares static equilibria and assumes well-characterized systems capable of manipulation 'from outside'." (Walker *et al*, 2001)

Each subsequent movement brings not only a new set of *policies* to be implemented, but also new tools and methods that are required to advance that cause and support the prevailing interpretation. An overview of these tools is set out below, but this research argues that there is no single best way.

Making explicit the competing methodologies, frameworks and rule sets for overlapping areas of governance and taking an adaptive approach that is context-sensitive and dependent in its approaches and responses, is a more appropriate way. In particular, surveillance is an increasingly common and highly visible emanation of these overlapping areas of governance. This is not to suggest another 'school' of thought, one that again brings its own tools and methods, but rather to offer a meta-approach to public policy, which takes into account the swirling, nascent policy environment, and consider where, when and how the policy world should crystallize into action and lived experience.

In assessing these approaches to policy, Ostrom lays out three distinct levels of specificity: frameworks, theories and models. Frameworks identify where more systematic analysis is needed, and which elements may be required; theories help us to understand which elements are relevant to specific questions; while models *"make precise assumptions about a limited set of parameters and variables."* (Ostrom and Ostrom, 2011:119).

Stages Model

This approach posits that policy-making proceeds through a series of sequential and chronological stages. Due to its straightforwardness, it is still commonly used, both to and by, newcomers to the policy field to *"impose some conceptual order"* (John, 2012:22) in a complex environment.

John offers the Clean Air Act 1956 as an example of when this model has operated as theorised, but notes that this type of example is all too rare; the simplicity of the model is its theoretical undoing, as it is incapable of carrying within it the interplay and dialogic nature of a more nuanced policy. Where there is disagreement, compromise, and negotiation, the policy will move in fits and starts, backwards and forwards through this process, which removes the linearity. John notes that writers will introduce loops (John 2012:26-27), but this does not overcome the main criticisms laid out above.

Lasswell (1951) was one of the first to approach the policy process in terms of phases, suggesting that policies moved through the following stages:intelligence, promotion, prescription, invocation, application, termination and appraisal.

Lasswell noted that the stages model was *"a conceptual map that must provide a guide to obtaining a generalized image of the major phases of any collective act."* (Lasswell, 1951:28). It is a useful heuristic, and has been greatly expanded on. Most notably Dror (1986:164) identifies three main stages: meta-policy-making, policy-making, and post-policy-making. Each of these contain sub-phases, providing eighteen in total, going some way to recognising the conflict between a rough heuristic and the detail that is sometimes required in the policy process.

Nakamura makes a similar criticism, arguing that the simplicity of the stages model makes it unrealistic (Nakamura, 1987), while even early research noted that delineation between stages could be unclear (Lindblom, 1959). Jenkins-Smith and Sabatier argue strongly that that the stages model

(paraphrased unless in direct quotations) Is not a causal model,that does not allow for empirical testing of hypotheses. They go on to note that as a heuristic, it is descriptively inaccurate, and lacks detail due to its top-down nature. Lastly, they argue that the stages model *"fails to provide a good vehicle for integrating the roles of policy analysis and policy oriented learning throughout the public policy process."* (Jenkins-Smith and Sabatier, 1993:3-4)

Howlett offers another stages-based model, which mirrors common political rhetoric of 'how policy works', covering agenda setting, formulation, decision taking, implementation and evaluation (Howlett, 2009).

There are some positives for this approach, in that it also focuses on the individuals and organisations that are involved in – and who 'make' – policy at each stage. It also identifies how policy-makers are bounded within operational and organisational constraints (Rose and Davies, 1994), with individual policies simply becoming variants on an on-going theme. This can lead, though, to policy as its own cause (Wildavsky, 1979) with policy-makers being required to make policy as a means of self-justification: It is difficult to argue that one's own job is unnecessary and that the best course of action is to do nothing.

It is notable that each of these approaches begins with an initiation phase, covering very similar territory, the problematisation of an issue:

"*Policy derives from the interactions of public opinion, interests, elites and ideas which are then filtered and structured by the institutions that guide the measure through the political system." (John, 2012)*

Thissen and Walker (2013), and Enserink, Koppenjam and Mayer (2013: 13) offer updated and contemporary general 'rational decision-making' models of policy-making. Although these attempt to widen the reach of the process, they are still predominantly based on 'modelling' approaches, continuing the assumption that there is a 'best' solution, for any given point.

Implementation Effects on Policy

For policy to be effective it must be implemented, otherwise it remains inchoate. In the tools laid out above, implementation is seen as a step in the process, and it is often carried out by different organisations from those making or taking the decisions[2]:

"*Policy implementation encompasses those actions by public and private individuals or groups that are directed at the achievement of objectives set forth in policy decisions*." *(*Paudel, N.R.*, 2009)*

Pressman and Wildavsky (1973) felt that each additional stage towards implementing a policy reduced the chances of correct implementation, and that this could be modelled numerically. An increasing distance between policy intent and implementation having a correlation to the complexity of the policy and its implementation does have some intrinsic appeal.

However, their model of success probability as 1*(0.9)^n means that the probability quickly drops below 50% for most policies when considering the succession of organisational layers policies will travel through to the ' front line'. This is consistent with Dror's view (Dror, 1986) that most 'problems' could be solved by providing a decision-making framework for people lower down the organisational ladder, with repetitive decisions within that framework taking up the bulk of the strain. Differing 'layers' of implementation will have differential impacts on the likelihood of success; a policy that is strongly supported within the political and policy sections of the UK government may still fail in the face of strong opposition, or even indifference from the vast numbers of staff who implement the day-to-day decisions of the policy framework.

Even if policies are created in a rational manner, the implementation is subject to a number of factors such as intra-organizational relationships, organizational capacity and institutional complexity (McLaughlin, 1987:12) that

2

affect the manner, speed and likelihood of implementation. Taking a more comprehensive view of the issues, rather than a clearly expressed bounded rationality, should be reserved for more complex issues of policy.

John notes that policy drift is a problem during implementation, and is about more than the *"...obstruction of policy goals by bureaucrats, but is the natural evolution of policy..."* (John, 2012:30). However, it is unclear whether this is a desirable method of policy evolution, or what the normative position with regard to this should be for the policy maker. If the implementation of a policy 'drifts' and becomes more or less effective, how should that be treated? And at an organizational level, how can the implementers of policy understand the difference between 'positive' drift and potentially harmful changes, and between the practical considerations of implementing a policy in reality and organizational resistance?

Walker *et al* (2001) see these questions as potential positives, and argue that changes and adaptability – including methods of feedback to understand them – should be built into policy from the start, acknowledging that it is part of a larger, recognizable process that 'traditional' policy sometimes obfuscates or avoids.

The issues around implementation and policy drift start to lead towards a less sequential or phased system. John noted the introduction of loops, but these are often improperly managed, and fail to iterate. As each layer of implementation applies its own context and creativity ('drift by interpretation', Kress and Koehler, 2005), the need for smaller, more reflexive loops of policy-making and implementation becomes more evident; this represents a fundamental change in how these needs are conceptualised, as policy attempts to deal with an accelerated and hyper-connected environment.

The Changing Project of Policy Analysis and Rational Choice Theory

A key approach in economics that is increasingly found in political science and policy analysis is the assumption of rational choice: individuals desire more of a 'good' (whether literal consumer goods or a benefit). Becker was awarded a Nobel Prize for Economic Sciences for his contributions to the field, particularly in finding rationality in areas where research previously indicated people acted irrationally, specifically criminality (Becker, 1968) and human fertility (Becker, Murphy and Tamura 1960). Rationality includes bounded choice, where individuals have a defined range of options (and abilities to decide), rather than complete freedom (Simon, 1972). Accordingly, people will weight costs and values *within this range* in making their decision; the implication for policy-makers is that they can create frameworks that allow people acting in the desired manner to maximise their outcomes.

The policy process (often used synonymously with 'legislative' process[3]), as a macro-framework for these decisions, can be seen as similar in pursuit of rationality, and simply seeks to select the 'best' solution. Maass (1966) posited the principle that the main bounding axis for policy concerns should be efficiency; this assumption became significantly more prevalent after the rise of NPM. Maas argues that the process itself should have the scope to make the necessary trade-offs:

"If the subject is highways, or most other public investment programs, perfect the efficiency benefit-cost technique for your agency's program. Once this is done there should be no difficulty in deriving through the legislative process a trade off between [economic] efficiency and another objective." (Ibid:225)

This view arises from logical analyses of risk-based games, rather than a psychological or social viewpoint that understands the motivations that individuals have (Tversky and Kahneman, 1986). This, and the lack of lived

3

experience in the analysis arguably limits its use in the complex environs of policymaking. Although the objective in the above example may be to build a highway, the economic efficiency and objectives being traded off did not necessarily take into account the quality of life of those near the roads, or environmental impacts. Although these can be costed and built into models of economic efficiency, as can behaviour (Kahneman and Tversky, 1979), the principle of economic efficiency and rationality is still sacrosanct. Sen notes that even now "...*benefits and costs have claims to our attention.*" (2000:934); the provision of economic and financial information provides a veneer of respectability implying a high degree of rationality. It offers faux legitimacy even where un-validated, and assumes an equality of value across all recipients.

Policymakers must assume both that the 'good' can and must be maximised, or that there is a minimum appropriate level above which the least expensive option should be sought: *satisficing.* The normative question of '*how are the sets [from which options gains and benefits are created and drawn] themselves created and drawn?'* lacks any engagement with the socio-organisational context that it exists within (Michalos, 1973). It provides little grounding on which to determine if maximisation or satisficing is the appropriate response: fundamentally, whether it is 'cheap', and better than 'good'.

Parker et al (2007:348) note that a constant pursuit of maximisation results in consistently poorer outcomes for individuals, and lower levels of overall satisfaction, regardless of the value of the outcome. The more rational a policy-making process attempts to be, the less suited to decision-making it is, particularly in high-stakes policymaking (Allison, 1971).

Majone (1989) argues that policy analysis adds to decision-making capability by challenging 'conventional wisdom'. Hogwood and Gunn (1984) noted similarly that it could help to define the problem. Both of these approaches are built on a fundamentally flawed premise of rationality, that the 'best' solution or problematisation will 'win'. Policy analysis and politics are littered

with examples of evidence-based policy that have been discarded as politically inconvenient (Rogowski, 2013). Cultural issues have taken precedence over economics on immigration (Hainmueller and Hiscox, 2010), welfare (Lupu and Pontsson, 2011) and voting systems (Iversen and Soskice, 2006). Green and Shapiro (1996) note at length that many studies which support Rational Choice Theory (RCT) are statistically weak (referring back to Sen's argument about faux legitimacy of economic points, even where incorrect). When these statistical weaknesses are corrected, the thesis fails; RCT features disproportionately in both economics and political science, when adjusted for its theoretical soundness, reflecting the 'hegemony' of modern economics in policy (Foley, 2003). Although the intent here is not to single out RCT approaches, it is clear that there is a disproportionate impact from this school of thought, which should be reconsidered, if not yet entirely set aside. John perhaps puts it most succinctly, reflecting that that '*[t]here is no inherent superiority of rational ideas*' (John, 2012:32).

Rationality and Incremental Decision Points

Applying these models, in spite of their severe limitations, the 'correct' approach may be to break with historic practice and policy (Lindblom, 1968) of assuming that policy-making organisations do so in a neutral manner. For these heuristics to be usable, it is necessary to assume that the policy actor is a single rational actor. However, in considering the interplay of, for example, government agencies interacting to create policy, an analysis quickly arrives at multiple actors. In isolation, this is an uncontroversial position: if the assumption that actors will continue to behave rationally is retained, then a tendency for incremental changes from the status quo is typical, and is arrived at through high volumes of negotiations (*ibid*). The lack of a single identifiable decision point means that incrementalism is often disjointed (David and Lindblom, 1963), moving us further away from the rational decision model, with a focus on risk avoidance and 'muddling through' (Lindblom, 1959).

Decision-making is contextual and influenced by location, time, and prevailing political environments. The level of congruence between potential policy solutions and existing political frames is given insufficient consideration in stages-type models. Etzioni (1967) proposed a 'mixed-scanning' model where the range of problems and available solutions is selected as the 'operating set' for policy decisions, making some of the bounds of rationality for the problem explicit. However, the range of available solutions is still heavily influenced by organisational context and influence.

The Art of Judgement – Policy as Relationships

Vickers (1965) saw the policy-process as always on-going, and as a set of rational relationships to be maintained, rather than the actions of a '*rational, purpose-ridden man'* (Vickers, 1965:45). Vickers' approach combined the components of a stages-like model into the objective of creating 'appreciative fields', which the well-intentioned and educated policy-maker can use to:

"regulate the relationship at some level more acceptable to those concerned than the inherent situation of the situation would otherwise provide." (ibid:43)

Vickers makes the case for policy-makers processing the subject at hand through cognitive schema and issue frames. Given Vickers' background within the comparatively conservative UK Civil Service, this takes a surprising, relativist viewpoint that argues that policies cannot be objectively judged as successful, and are only 'approved' or 'condemned' by the judgements of others. Vickers' position can be interpreted as stating that policy-as-a-response and the situations that elicit responses are both socially constructed. Long consultation and negotiation periods that precede decision-making are part of that construction (Loveridge, 1979).

However, Vickers does make a fundamental assumption that the intent of the policy is visible; this model's own underpinning and implicit assumption of rationality, based on Vickers' experience of the 'gentleman's agreement' approach that characterises much of the book. McConnell notes the plethora

of reasons that a policymaker may choose to conceal both intent and outcomes of a potential policy, ranging from the political and ideological to simple tactical choices (McConnell, 2010:91). Policy as a relationship significantly downplays the importance of institutions, which is particularly curious for a UK Civil Servant. Service roles have been retained even as individuals are moved between them since the Northcote-Trevelyan report (House of Commons, 1854). This movement of individuals between roles makes the relationship between individuals and organisations more important. Even if this were considered, it still conceives of policy as a solely elite activity governed by insider individuals and organisations, downplaying the effects of media, discourse and the public.

Institutionalism

Institutionalism is a sociological approach to organization and political theory, which argues that institutions are the main actors in society and politics:

"It is necessary for the relations within the structure [of the organization] to be determined in such a way that individuals will be interchangeable and the organization will thus be free of dependence on personal qualities. In this way, the formal structure becomes subject to calculable manipulation, ***an instrument of rational action."*** (emphasis added) (Selznick, 1948: 25)

The primary role of institutions, if not their purpose, is to provide stability to the policy, a degree of organisational memory. Early institutionalism focused on formal institutions: government departments, agencies, and units within those organisations. This formalisation and memory, including the capturing and crystallisation of rules, mean that there is an understandable platform and system for those engaging in policymaking. Although Selznick noted that institutions can and did have dichotomous interests, these early institutionalist approaches have largely been discarded due to their limitations and narrow focus.

"In politics as in everything else it makes a great deal of difference whose games we play. The rules of the game determine the requirements of success." (Schattschneider, 1975:48).

Considine (2005) points out that policy institutions can be both an impediment or barrier to action, as well as a force for creation. Considine focuses on institutions because:

"much of policy-making is embedded in the routine practices carried out by government agencies employing well-worn repertoires of action."

However, much of what Considine considers here could more properly be regarded under the frame of implementation theory. There is a significant difficulty in isolating when a policy 'becomes' a policy. It is necessary to understand when and where the policy process realises decision or inflection points, as distinct from the point it becomes a routinised application or implementation of a previous decision. Such delineation is poorly served by much of the literature on public policy, as it fails – in many instances – to take into account such complexity and ambiguity in a systematic or replicable manner.

New Institutionalism

New institutionalism rejects the formal institutional focus and rational-actor assumptions that underpin classical institutionalism. Instead, it considers institutions as contested and constructed, and as actors within a context shared with other actors. An example of shared context could be a change of Government (Karl, 1990), where the context shapes the process. The requirements for stability, expressed through the desires and responses of governmental agencies overrides even the electoral; manifestos on which deregulating governments (in this example) were elected.

Both change and stability can be explained through New Institutionalism (Immergut, 1992; Hall, 1992), reflecting how the 'new school' is not a cohesive

or unified theory, but rather a collection of theories valuing and engaging institutions with varying approaches to construction and interaction.

Steinmo (1993) offered institutions as the main driver of tax policy variances between Sweden, the UK and the USA. The fragmented nature of the US federal systems atop a range of states with varying policy environments means a national sales tax has never been introduced. Sweden, with a single national body – and a proportional representation voting system that provides stability – means that the ruling coalition could pass such a measure. The role of state-level governments in the USA is given insufficient weight, as is the interaction between state fiscal policy and the federal government, with several areas of the US having both state and county (or city) sales tax variances. His findings are also inconsistent with studies (see, e.g. Martin and Vanberg, 2004) that show 'European' politics, driven by coalition and compromise systems, tend towards moderacy.

Hall (1992) offers one of the few institutional analyses where institutions account for *change,* rather than stability. Hall examines the Conservative government's monetary policy change from Keynesianism to Monetarism. It examines how formerly oppositional think tanks were created and came into power alongside the party, with the effect of policy on the financial markets taking centre stage on analyses and decision-making. John (2012) strongly argues that this scenario might be as much about the realignment of political forces and alliances than just the institutions. The institutional environment was complemented and altered by the introduction of new institutions and advisers, a situation that was repeated after the 1997 and 2010 elections.

There are three main strands of criticism of New Institutionalism. The first is that comparative studies often assume a consistent starting base between countries (as in Steinmo), where the societal context may be very different. The organisations responsible for social security and welfare policy in the UK and US differ significantly, and have very different remits. Indeed, the very meaning of social security is quite drastically different in the two countries. Secondly, there is a degree of assumption around similarities between policy

departments within the same country. Where studies have looked at, for example, the Department for Work and Pensions (DaGuerre, 2004), they may not be generalizable to the Home Office or other departments. Because of the disparate nature of departmental remits, and the different ecosystems of actors around them, comparative studies across departments should note the differing contexts. Finally, the very definition of institutional is too variable, ranging from very narrow – as was the case with traditional institutionalism – to very broad, allowing even very tangential organisations the designation and weight of an 'institution'.

Institutionalism is a useful but limited approach, for reasons of applicability, generalisability, complexity and practicality. What would be desirable, and is addressed later, is an approach that can take on the contextual values of new institutionalism, without potentially isolating the research in a particular, singular context.

Ancillary Approaches

In proposing their general 'rational decision-making' model of policymaking as noted above, Enserink, Koppenjan and Mayer (2013) outline the heuristic models that are available to the policymaker and analyst:

Rational decision-making process

Political Game

Policy as discourse

Garbage Can model

Institutional process

The first and last have already been examined, but the remaining models offer potential new ground.

Policy as a Game

Here the policymaker's goal is to implement policies *because they score a political victory for the government, or over their opposition.* Being led by political values, driven by party ideology, positive outcomes are of secondary importance, and this is particularly common in two-party systems (Alesina *et al*, 2001). Institutions as well as individual politicians play the game (Immergut, 1992), and the nature of the game is highly contextual; participants triage and consider such positions as whether this is a repeated game, whether winning will damage their chances in future rounds, and whether a particular outcome damages the equilibrium of the game.

The determinant of 'winning' policy centres on outcomes for game participants, not for policy users. Media coverage tends towards the mechanics of the game, as this lends itself to adversarial and conflict-based media frames. Media coverage is far more likely to report policy-as-a-game when it has national impact, or where there is significant dissonance between the government and opposition (Lawrence, 2000). They are equally unlikely to report on systemic issues, which exclude minorities and those outwith the policy community from the game (Bachrach and Baratz, 1970). Policy-as-a-game is subjective and heavily mediated, which relates it closely to policy-as-discourse.

Policy as Discourse

Although discourse defies a single descriptor (Bové, 1990:53), it is helpful to understand the term in the context of policy. Policy as discourse encompasses spoken or written interactions between individuals that creates, shares or contests meaning. From this, Bacchi argues:

"The premise behind a policy-as-discourse approach is that it is inappropriate to see governments as responding to 'problems' that exist 'out there' in the

community. Rather 'problems' are 'created' or 'given shape' in the very policy proposals that are offered as 'responses'." (Bacchi, 2000)

Fischer and Forester (1993) argue that policy-as-discourse is theoretically sound. Being built on the public sphere concept (Habermas, 1989), policy as discourse places rational debate at the heart of the process, with policy being formed through exchanges (Dryzek, 1993), arguments (William, 1993) and shared meaning generated through those processes. The quality of policy (and sustainability of the model) is dependent on the quality of discourse.

Participants in the policy process retain their own belief systems and values (Sabatier, 1987), which layer with organisational and role-based systems and values to create policy paradigms (Hall, 1992:25). Assuming discourse is formed of dialogue and utterances that interact with each other rather than the longer cycle of rhetorical utterances in the press, then the iterative shaping of policy discourse becomes clearer. Shaping happens within the policy community, producing a set of shared assumptions and rules that both create meaning on an on-going basis, and also bar outsiders from the process.

Political environments in the UK and United States have been increasingly polarizing (Iyengar and Hahn, 2009; Flanagin and Metzger, 2017). Value clashes – which inform the cognitive schema and policy paradigms that actors possess – have increasingly been the only interaction between divergent groups, as groups become progressively more isolated from each other. Fischer and Forrester posit these as 'dialogues of the deaf', where there are:

"...enduring impasses about substantive views, in which parties talk past each other using reasonings (sic) that are plausible by themselves but mutually exclusive." (Fischer and Forester, 1993)

In media framing terminology, engaged dialogue occurs where discursive frames overlap or converge. Here the issue is frame divergence, where opposing frames come into contact with decreasing regularity, so the core beliefs that underpin them go unchallenged, become more entrenched, and

more difficult to change in either direction. The policy-as-discourse concept shares more of the negative aspects of the 'echo chamber' (Jamieson and Cappella, 2008; Wallsten, 2005) than of the ideal type prescribed by Fischer and Forester, or the ideal of the Habermasian public sphere.

Discourse should be considered an expressive element of what is later described as the 'policy assemblage', explored in Chapter Five. Self evidently, discourse is also closely linked to issues of media framing, as much policy discourse is mediated through news outlets (Entman, 2003), and so the operations and machinations of media framing take on an additional level of importance for policy-as-discourse.

Garbage Can Policy

This position posits that policy is primarily a response to disruption and coincidence. Kingdon argues that "*A decision situation is like a garbage can into which participants deposit all sorts of problems and solutions.*" (Kingdon, 1995). Opportunities to control the process are essentially limited by the resources available at the time, which brings parallels with Etzioni's mixed-scanning (Etzioni, 1967) – identifying a set of available problematisations and working with those. In a departure from the cognitive frameworks of earlier models, this posits a conscious move away from unconditional rationality.

Policy 'windows' may be opened by a crisis, the cessation of an activity, or by a simple confluence of several streams of activity. This model is significant as it is one of the first to acknowledge the limitations of rationality in the process. It contrasts with Lindblom's 'muddling through', as there is no inherent position of risk avoidance, but rather each problematisation is dealt with at the time of occurence with whatever resources – political orfiscal – are available at that point. This concept of congruence around a particular issue, scenario and point in time is crucial, and is explored further below.

Kingdon (1995) extrapolated from the Garbage Can model, and offered that there were three streams within policy-making: politics, participants and problems. Streams can be harnessed at opportune times, leading to the rise of decision-making agendas; Kingdon also argued that it was 'policy entrepreneurs', individuals who are willing to invest their personal time and resources into projects, who can bring these streams together and 'make' policy happen. Kingdon's model is widely influential, and is a significant entry point into discussing the availability of elements of a decision-making model, and how these can be made available. However, there are some criticisms of Kingdon's model, in the first instance that it fails to adequately provide tools for mesoanalysis[4] (Bundgaard and Vrangbaek, 2007), while Exworthy and Powell's (2004) posit that viewing the streams themselves is insufficient, and that analysis should instead seek to understand the content and creation of the streams present within policy-making.

The Mobilisation of Bias

Dahl asserted that the existence of a policy elite:

"can be strictly tested only if… there is a fair sample of cases involving key political decisions in which the preferences of the hypothetical ruling elite run counter to those of any other likely group that might be suggested." (1961:466)

However, this is embedded in a deeply rational model, in a similar manner to earlier criticisms. It depends on a deeply held assumption that policy actors' motives, desires and actions are entirely visible and openly contested. Bachrach and Baratz also note that power is typically viewed as only being manifested in concrete decisions. However, for many groups, particularly the marginalised, the effects of power are often felt in non-decisions, through the '*mobilization of bias*' (Bachrach and Baratz, 1970:11). 'Non-decisions' are the means by which "*demands for change… can be suffocated… or killed before*

4

they gain access to the relevant decision-making arena." There is a clear resonance with agenda-setting in media framing models, which is rarely explicated in policy analysis, where the media:

"may not be successful much of the time in telling people what to think, but it is stunningly successful in telling its readers what to think about. The world will look different to different people… depending on the map that is drawn for them by writers, editors, and publishers of the paper they read." (Cohen, 1963)

Bachrach and Baratz go on to note that "*policy choices are frequently made in the absence of a clear-cut, once-for-all decisions. They simply 'happen', in the sense that certain steps are taken that are necessary."* (*op. cit*:42).

In this theorisation of the policy flow, mobilisation tends towards the status quo and occurs around conflicts of values. Consider a potential policy change, which conflicts with majority values. Those values, and threats towards them, are the mobilising factor for the current dominant group, as well as being a source or result of their existing authority, influence or power. In coalescing around those values, the dominant group creates barriers to entry for minority opinions or changes to the status quo, which take the forms of:

Non-decisions by exertion of influence through community values, or procedures and institutions

'Defeating' the conflict through decisions by institutions

Frustration through ineffective implementation of a policy

The mobilisation of bias offers a chance to capture how unlikely alliances and coalitions of common self-interest come into play to perpetuate the status quo. These alliances are not necessarily rational in any conventional sense, nor are they perpetuated in a consistent manner, being called into being only when needed. The degree of agency involved in 'being called' can easily be overstated, and care should be taken to not over-assign agency in the

creation of a relationship supported and drawn together by mutual, overlapping self-interest. This has a significant resonance to the new approach of conceptualising public policy discussed later, and it also more closely aligns the modernity of politics and subjectivity of those outside the policy 'bubble'.

Policy Success Heuristics

As already noted, policy processes and their descriptors can be regarded as heuristics for a complex and multi-dimensional process. As the notion of success is often contested and contextual, McConnell (2010) writes extensively on both the political problems of defining policy success, and the need to encompass that complexity when measuring success. The concept of success is important in any policy model or analysis, as it frames debate, and outputs from one policy process arguably constitute an input into the next, regardless of whether the conceptual models are capable of accurately capturing, portraying or responding to them.

Policy success heuristics can be grouped in three main areas, with McConnell's success heuristics noted:

Rational, apolitical analysis	Cherished values – inclusion and responsiveness	Executive centric & Technocratic
Process Success	*Political Success*	*Programme Success*
Stokey & Zechauser (1978) – societal well-being depends on intervention to correct market failures	Schneider and Ingram (1997:203) – high level moral and democratic values	Daly *et al*(2006) – Policy priorities established and used as metric
Mitraney (1996) – economic and technical cooperation is	Lindblom (1959, 1965) – negotiation and bargaining	Cohen et al (2008) – long term

gradualist approach to peace and stability	as good results in their own right	professional management of policy by objectives
	Gastil (2008) – policymaking requires inclusion and deliberation (cf. Habermas)	

In turn, each of these success heuristics can be durable, conflicted/contested, or precarious:

"Overall, one of the lessons for present purposes is that when policy analysts and commentators go in search of 'what works', they come up with widely different answers." (McConnell, 2010:81)

This re-demonstrates the conditional and mixed rationality at play, creating a conflict between 'real' and 'constructed' definitions of policy success. The UK Oil Blockade in late 2001 showed different definitions of success for the Government across the time of the blockade, from the upholding of 'law and order' at the beginning, to simply ending the blockade; in turn these definitions of success differed from the public's, which sought to influence taxation policy downwards (Robinson, 2003).

Policymakers and politicians must understand the influences on them at any given time, in terms of institutions, actors, media coverage, and how that *coalesces* around a particular policy. The presence of particular elements, as well as their direction and strength, varies over time and topics: the goalposts for success are always in flux.

Summary

A significant disconnect emerges from the literature, between the general theoretical frameworks for policy-making, and the lived experience of the process, particularly as concerned by those who are subject to the flows of power from those processes. While the policy process has traditionally been viewed as rational (or contained within a set of commonly shared bounds of rationality), it increasingly struggles to deal with, and account for, complexity, pace, and the fragmentation of media. The desire to have a demonstrable policy on *anything,* necessarily leads to policy frameworks and machinery of government that must have a policy of *everything.*

Rational models require knowledge, particularly meta-knowledge of what is knowable or not in any given situation. Foucault appositely captured the self-perpetuating policy machine when he noted that:

"My point is not that everything is bad, but that everything is dangerous, which is not exactly the same as bad. If everything is dangerous, then we always have something to do." (Foucault, 1997:256)

The divergence between the likelihood of crime or terrorism is vastly disproportionate to the amount of policy energy that these issues consume, particularly given the focus on treating the emanations of these issues, rather than their root cause. Politicians are equally unwilling to accept risk as they are views that are incongruent with their ideology. The rise of a risk culture that examines not just what is *probable,* but what is at all possible (Füredi, 2009) brings out the worst in these rational models, requiring ever-greater amounts of evidence, power, influence and authority, which exposes further risk that must be exposed, all the while in conflict with the ideological party systems (Rogowski, 2013).

"There are no 'optimal' or 'best' solutions: only politically negotiated, acceptable and feasible solutions…'technical scientific rationality' must accommodate to 'political rationality'" (Enserink et al, 2013).

The political arena is an almost entirely socially constructed one; it is entirely a product of the sets of relationships and customs that have accreted over time. Even within the pseudo-rational models described above, the institutions' chosen to carry out analyses are still the product of societal processes.

Within the rational models outlined here, evidence and empiricism is reified, except where it is politically inconvenient; the arguable cognitive effect of that on-going reification of evidence is that evidence and empiricism are assumed, even where they are not present. If this rationalism falls to the wayside, then it is necessary to consider another method that incorporates and weights (or at least adequately acknowledges the interaction of) those aspects of the socially-constructed against evidence and process.

Policy – whether outcome or process – should be seen as the coming together of a set of interests and movements, politically, discursively and bureaucratically, which has an effect; an output. Rather than a defined process, it is a conglomeration of self-interest, disinterest and narrative convenience. However, that collection of conflicting rationalities and competing actors encompasses the set of desires and possibilities for the governance of the modern state. As a result, considering and reconciling the relationships between public action, public policy and governmentality becomes essential.

Chapter 3:

Reconciling Governmentality and Public Policy

Introduction

Governmentality, as now used refers to the "*study of the rationalisation of governmental practice in the exercise of political sovereignty*" (Foucault, 2007:4). Although initially posited in a series of lectures that were never directly published, it has received considerable critical attention and use as a theoretical framework, with Dean defining the concept as:

> *"Government is any more or less calculated and rational activity, undertaken by a multiplicity of authorities and agencies, employing a variety of techniques and forms of knowledge, that seeks to shape conduct by working through our desires, aspirations, interests and beliefs, for definite but shifting ends, and with a diverse set of relatively unpredictable consequences, effects and outcomes."* (Dean, 1999:11)

These techniques and forms of knowledge are mentalities, conditions of *"forms of thought… not readily amendable to be comprehended from within its own perspective."* (*ibid*, 11). This emphasises the recurrent theme from public policy, that the *"thought involved in practices of government is collective, and relatively taken for granted.*" (*ibid*, 16). Governmentality is offered as an alternative approach to the multiply layered issues of government, governance, politics and public policy issues that can:

> *"provide a theoretical elaboration which potentially opens everyday and institutional programmes and practices for critical and tactical thinking, it also provides a considerable array of empirical work in terms of which interventions can be examined and thought out."* (O'Malley, Weir and Shearing, 1997:503)

Being more descriptive and reflexive, governmentality offers less a definition of public policy processes, and more an approach or praxis towards understanding them. Governmentality attempts less to offer a model of how policy can or does work, and attempts to internalise, through the 'analytics of government' methods to describe and interpret how those processes *have* worked in the past, without necessarily making claims to repeatability. In contrast to O'Malley's viewpoint, this description and interpretation can occur without necessarily providing attention to, or analysis of, specific programmes and policies (Larner, 2000), even when these programmes or policies may be the resulting output.

Governmentality has two broad meanings. The first is simply how to think about governing, in collective terms; how the different mentalities of government are encompassed, as well as how its power relationships and structures have developed and are maintained (Dean, 1999:16). Dean goes on to argue that government, in his Foucauldian definition, has its own '*subjective modalities*' of desire, aspiration, interest and belief. The second is the emergence of a new way of thinking about government, the requirement that the population is controlled in non-force terms (what Foucault termed the '*micro-disciplines*'), and the subsequent emergence of a political economy around this new way of thinking.

In the models of public policy examined, there is an attempt to create a universal understanding of the policy sphere: a bounded and rational model that provides continuity and consistency, an episteme of constant policy. However, there is also an opportunity to understand the world through difference, through "*the divergence, the distances, the oppositions, the differences, the relations of its various scientific discourses."* (Foucault,

1991:55), particularly in the political and policy spheres. This can lead to a sharp delineation between *political* governance, in an ideological and party political sense, and Foucault's '*problematic of government'* (Foucault, 2007:89), encompassing the general issues of personal and societal conduct with which a government is bound up.

The state has the same qualities as politics and economics, which are:

> *"neither existing things nor illusions, errors or ideologies. They are something that did not exist and that is part of reality, [that is] the effect of a regime of truth that separates truth from falsity."* (Foucault, 1997:22)

Lemke considers that the state is a transactional reality (Foucault, 1997:31), layered from a "*dynamic ensemble of relations and syntheses that at the same time produces the institutional structure of the state and the knowledge of the state."* (Lemke, 2007). However, this multiplicity of relations is still predicated on the assumption of rationality, per Dean's definition, albeit the '*play between competing strategic rationalities'. (*Barnett, et. al., 2011)

It is questionable whether the interplay of a multiplicity of rationalities that are only internally coherent is, in fact, rational. Perhaps existing policy processes and models have served only as boxes that attempt to constrain the intrinsic incoherence and conflict of competing systems of belief and action. This particularly holds when considering the rationality of abstract policy versus the sometimes absurd and irrational lived experience of policy for individuals.

Despite conceptualising the dual genealogies of the modern state and the modern citizen, there are few sites for analysing the nature of the relationship between the two. Lemke allows that the governmental approach, acknowledging the ongoing process of policymaking and state formation, allows the observer's position to exist with theory construction (op. cit.,7). However, the existence and creation of state structures provides a symbolic infrastructure for sites of (high level)

intervention, and provides a framework within which state subjects live their relation to the state, but without necessarily critically assessing that relationship.

Practical Governmentality and Public Policy

Governmentality and public policy diverge when the relative 'outputs' of the two are considered. Although policy is focused on specific outputs (which may be behaviours or conduct), governmentality considers the act of governing as more to do with governing the range of conducts available. Governmentality considers how the '*emergence and stability of state agencies is intimately tied to the incessant generation, circulation and repression of knowledge*.' (Lemke, 2011:28) The state and policy are not the same, nor are they causal factors that emanate from each other, but they are mutually constitutive and reconstruct each other at every inflection.

Discussions of policy processes focus on the manifestations of power without necessarily explicitly critically examining them, and the underlying technos for the approach is taken for granted in most instances. Governmentality displaces these traditional assumptions in three ways:

- Considers power beyond the consensus/violence dynamic
- Differentiates between power and domination
- Considers and clarifies the relationship between ethics and politics

While public policy analysis is primarily concerned with the means by which policy is achieved (and therefore the means by which power can be accrued, exerted and maintained), governmentality is concerned with the *techne*[5] of power and control:

> *"...by what means, mechanisms, procedures, instruments, tactics, techniques, technologies and vocabularies is authority constituted and rule accomplished?"* (Dean, 1999).

5

In contrast to other models of public policy, which are concerned with the specific actions that governed populations may take, governmentality is more concerned with how the range of options available to a population is shaped (Lemke, 2011:18). As this may often occur implicitly, and without the knowledge of those so restricted, it differs from public policy models – particularly the more rational or deliberative models – which begin with an impetus towards an explicitly discussed change in behaviour.

While Foucault, Lemke and Burchell (*inter alia*) use power in a broad sense, Bachrach and Baratz (1970) made the case significantly earlier that authority, influence and force should be delineated. Each represents a different relationship of mind between two or more parties, and this distinction is crucial to critical analysis of governmentality and public policy.

- Power: a conflict of interests or values where one party bends to the other's will *without the use of sanctions;*
- Force: a conflict of interests where one party 'makes' the other comply, removing the option of voluntary compliance[6];
- Authority: communication with qualities such that the recipient obeys the instruction; and
- Influence: as power, without the tacit or overt threat of sanctions.

Although compelling, these delineations may not be as clear-cut as initially presented. The overlaps between the categories and the degrees of co-construction between, for example, authority and power, or between influence and force are open to a wide degree of interpretation. Despite presenting influence as power without the threat of sanction, there is a strong case to be made that all influence extends, to some degree from either self-interest, or from implied risk of sanction.

Lemke is particularly strong in noting the scope and limitations of governmentality, particularly that '*it is mostly the territorially sovereign nation*

6

state that serves as the implicit or explicit frame of reference.' (Lemke, 2011:4).

Dean (1999) goes on to note that:

> *"An analytics of a particular regime of practices … seeks to identify the emergence of that regime, examine the multiple sources of the elements that constitute it, and follow the diverse processes and relations by which these elements reassembled into relatively stable forms of organization and institutional practice ..."*

Dean's conceptions of regimes of practice here are arguably interchangeable with the concept of policy. The obvious question for analysts is 'why is this approach not used in *considering* policy? There are a number of potential reasons, one of which is the regime of practice within the policy establishment itself: the ultimate purpose of any such establishment is the generation and analysis of policy. In many instances, examining the constitutive sources of that institute would arguably result in the removal, or at least fundamental change of the policy institution. Further, the issue of 'not reinventing the wheel' becomes a dominant practice that discourages each level of reflexivity: the apparatus of government needs to be able to consider itself - and the environment within which it operates - as broadly meta stable.

This seminal interpretation of governmentality focuses too heavily on macro-level considerations. Dean addressed this in later editions of the same text, noting that he had intended an analytics of government as a tool of criticism ('*open, multiple and immanent'*) rather than critique, *'conducted under universal norms and truths and pointing towards a necessary end.*' (Dean 2010:3). Rather, from the perspective of understanding policy and providing critical insight, the critical gaze should fall not just on the macro-level, on the sweep of general movement and discourse, but also on the relationship between lived subjectivity of individuals and back to the whole.

Lemke considers that "...*by focusing on direct and distinct technologies, an analytics of government avoids the pre-analytical distinction between micro- and macro-level, individual and state...*" (Lemke, 2011) and goes on to argue that this approach "*...makes it possible to ask questions about the relationships between different governmental technologies."* Where Lemke seeks to reposition and consider state-level issues, this research seeks to understand the relationships that are constituted between assemblages of the state and policy, and those that are subject to them.

In this vein, Dean's notion that "*Government as the 'conduct of conduct' entails the idea that the one governed is, at least in some rudimentary sense, an actor, and therefore a locus of freedom*." (Dean, 1999:15), is problematic, as it assigns a systematic level of agency to actors through their lived experience.)

In particular, Dean goes on to note that:

> *"Practices of government cannot be understood as expressions of a particular principle, as reducible to a particular set of relations, or as referring to a single set of problems and functions… rather they should be approached as composed of heterogeneous elements having diverse historical trajectories, as polymorphous in their internal and external relations, and as bearing upon a multiple and wide range of problems and issues."* (Dean, 1999:29)

Emerging from Dean's analysis is the concept of differing and divergent historical trajectories, a concept that is returned to in Chapter Five, when theory turns to understanding how practices of government are created and replicated. Key to his analysis is the notion that these practices of government cannot be seen as replicable applications of generalised principles, or reduced to a simple set of rules or organisational frameworks. Instead, the governmental arrangements and systems that we have in place

today are products of their own histories, and the series of problems and policy issues that they have interfaced with over a significant period of time.

Dean succinctly sets out the greatest strengths – and weaknesses – of governmentality and an analytics of government:

> *"..an analytics of government marks out a space to ask questions about government, authority and power, without attempting to formulate a set of general principles by which various forms of the 'conduct of conduct' could be reformed. The point of doing this, however, is not to constitute a 'value-neutral' social science. Rather it is to practice a form of criticism… that seeks to make explicit the thought that ... is largely tacit in the way in which we govern and are governed, and in the language, practices and techniques by which we do so. By making explicit the forms of rationality and thought that inhere in regimes of practice…an analytics of government can remove the taken-for granted character of these practices."*

Governmentality allows us to retrospectively understand the roots of a specific problem. This includes the genealogy and heritage of our current problematisation, as well as explicitly understand how historical and cultural accretion have allowed a comparatively stable system that is reasonably predictable.

For public policy – the creation, administration and subjectivity of public policy – it leaves the theorist at crossroads. What can be done or changed in public policy as a result of that seeming stable and predictable system? At this junction between understanding the generalities of the approach and taking or making action as individuals or policymakers, the key issue is the *operationalisation of* governmentality. As a method of analysis it is useful to consider how it can be used to understand and improve the relationship between pseudo rational policy process and a multiplicity of lived and chaotic experiences.

Normative Governmentality and Ethics

In addition to these concerns, Dean notes the presence of another layer that should be considered: that of '*what constitutes good virtuous, appropriate, responsible conduct of individuals and collectives.'* (Dean, 2010). Dean's conception of the 'conduct of conduct' attributes the governed individual a 'locus of freedom' (*ibid*:21-22). This views the individual as an actor with some agency, even in extreme circumstances of government, such as capital punishment or the exercise of torture. Individuals are free to operate within the bounds that the actions of government have put around them. What governmentality may inadequately capture are the effects of power-flows at the edges of the bounded space. Where the rationality around the governed subject boils over into irrationality (at least in terms of the policy system which governs it), the subject can be the object of significant acts of structural violence by the governing body.

In the example given of the prisoner on Death Row, the prisoner is, once again, free to act within the bounds of the governed space. This is – only and precisely – because he has been captured, processed through the juridical legal portions of the governed space, and is now in a more tightly striated space. The action that brought him there was not an act within the governed space's mentality, but rather a departure from it – an edge case that brings him into significant conflict. The now-prisoner has, in effect, discovered the limits of 'freedom' within the original governed space (wider society), and now finds themselves in a more tightly governed space with a much smaller locus of freedom (the prison and judicial system).

In the terminology introduced later, this action has become a line of flight that deterritorialised once-governed space, and transitioned the actor into another governed space, through an act of force. Governmentality is, in some respects, clumsy at handling the transitions between these spaces, in part because it views almost everything as inside the governed space; with that held to be true, every action is within a mentality of government. A key

counter-argument to this is that not all spaces are governed, and a focus on actors transitioning between governed spaces in governmentality quickly runs out of recognisable spaces between which to make such transitions.

Dean considers "t*he distinction between relations of power that are open… and those that are not is a useful analytical and descriptive tool. However, to the extent that an analytics of government endeavors to avoid global or radical project, such a distinction cannot be used to construct a general normative stance."* (Dean, 2010)

This may be the case if, as Lemke points out, the default frame of reference is the constituted nation-state. However, those power relations arguably also exist and cascade through every level of society and its structures. It is difficult to see how a principle that applies at state level may generate a normative principle applicable at interpersonal. Instead, analysis should take into account not just how an analytics of government can create such a normative stance, but also how that normative position recurses and modifies as it moves 'down' through the system.

The notion of open and closed relations of power is one that recurs throughout governmentality and the concept of liberty (or at least a sense thereof) being correlated with the openness of power dynamics is similar to the concepts of territorialisation and deterritorialisation[7] utilised in Chapter Five in terms of assemblages of government. Foucault argued that the more closed a relationship of power was, the more firmly set and 'congealed' it was: "*in such a state the practice of liberty does not exist, or exists only unilaterally or is extremely confined and limited.*" (Fornet-Betancourt *et al*, 1987).

Just as Foucault recognises the rigidity of some systems of closed power, systems that are strongly territorialised are tightly bound with, for example, the person-to-system relationship being highly structured, governed and controlled, with a high degree of power asymmetry.

7

These distinctions (of openness and power symmetry) are a helpful typology of the actuality of power relations, in terms of both understanding and theorizing public policy processes and systems of government. The typology also assists in conceptualising why coalescence and crystallization are particularly useful concepts. Within the surveillance studies context, differentiating between these types of relationships when examining a surveillance system or approach provides additional clarity, given the *de facto* removal of choice in the operation of a number of surveillance and social control systems.

Governmentality still indicates a degree of intentionality and agency over the exercise of control over citizenry. Foucault acknowledged the slow emergence of the problematic of government in response to the needs of a population-based state, but the approach retains a strong assumption around the degree of agency and rationality in how the government and its various mentalities operate. In effect, it has lost the effect of accretion and chance, of coincidences that occur, and impact the multiple rationalities that individuals and their managerial fiefdoms try to maintain.

The three levels that governmentality operates at can be mapped across to the public policy process more familiar to those with a management background. This attempts to bring together both approaches, to combine the replicability and comparatively low level unit of analysis that can be gained through management approaches, with the societal context and understanding of power flows that governmentality brings. In a number of cases, the overlap and congruence between the two approaches is such that all that separates them is terminological distinction, reflecting the differing disciplines.

Dean further suggests an analytics of government, characterised by four "*reciprocally conditioning yet relatively autonomous dimensions*" (*ibid, 33)*:

- The fields of visibility of government
- The techne, or technical aspects of government: the technologies and

mechanisms it uses

- The episteme, or forms of thought that are used in government
- The formation of identities: what forms of identity and self are assumed (and produced) by government?

Governmentality's 'more-or-less-systematised' middle-ground between violence and consent is where public policy analysis and process is fundamentally targeted. It attempts to understand, critique, and in some instances replicate or formulise the creation and exertion of power flows, but the distinction between directly compelling citizens to undertake a particular course of action and simply limiting their fields of action more generally is often lost.

Increasingly, this is skewed towards direct interventionism. Füredi's argument about risk management can be extended backwards towards the source (Füredi, 2009). Because it is possible that anything can happen to the capital systems and flows that support the western neo-liberal world, it is necessary to put in place everything that system requires, or could possibly be perceived to need.

In both instances though, the variability of both the process of governing, whether from a 'conduct of conducts' perspective, or from an attempt to systematise policy analysis (whether *for* or *of*) gives a degree of inconsistency between approaches; studies are highly dependent on contemporary context, a factor which is rarely captured.

Conclusions

Governmentality as a framework or theory is strong after the fact, and represents an almost archaeological approach, unsurprising given Foucault's focus over a number of years on the analysis the 'discursive traces of the past to write a history of the present.' Foucault viewed that he ''*... set out from a problem expressed in the terms current today and I try to work out its genealogy. Genealogy means that I begin my analysis from a question posed*

in the present' (in Kritzman, 1988: 262). This has two impacts in terms of using governmentality in and for policy analysis. The first that it is inherently weaker on future issues - governmentality makes fewer claims as to repeatability or future meanings that it derives from this genealogical process. The second is that those same issues – of multiple rationalities and a multiplicity of actors and levels – should be repositioned and examined from the perspective of the assemblage.

> "*heterogeneous elements with their own pre-histories are thereby reworked and readjusted to produce "phenomena of coagulation, support, reciprocal reinforcement, cohesion and integration*" (Foucault, 2008a, 239)

These phenomena of coagulation and reciprocal reinforcement that Foucault notes have constant readjustment and rework are exactly the terms that lead from considering Governmentality, towards theorising about systems of government as emergent and ongoing systems. Instead of focusing on the archaeologies of governments and governance past, it is necessary to theorise about complex systems that are continually performed and created, and how these systems cohere and perpetuate. In particular, it is necessary to understand how such systems flow from the macro level to the micro level of lived subjectivities. Chapter Five addresses how theory can move towards a reflexive and recursive system focused less on the attributes and components of systems, and instead how it can criticise and consider the relationships between those components, and the systemic qualities that emerge from such consideration.

Chapter 4:

Understanding and Theorising Surveillance

Introduction

This chapter outlines definitions of surveillance, and the consequences of those definitions within the context of surveillance studies and public policy. It examines theoretical approaches to surveillance, in order to understand the lenses through which surveillant systems and acts can be critiqued. In doing so, this section also attempts to take a critical view of what it means to research in surveillance studies.

Surveillance is a dominant and identifying feature of contemporary – and postmodern – society. It extends and replicates many of the structural imbalances and inequalities of late capitalism. Although widely accepted as either a consequence (Lyon, 2007) or an intrinsic feature (Giddens, 1991) of modern society, surveillance may not be inevitable. The co-building between surveillance mechanisms and economic systems (Fuchs, 2011) means that it may be disrupted as economic systems shift, or that resisting and disrupting surveillance mechanisms consequentially disrupts those economic systems.

Lastly, this chapter moves from the broad definition of surveillance drawn from the literature to providing, with a rationale, a subset and focus of surveillance that is examined within the bounds of the research. The field is simply too broad to try and operationalise any meaningful research across all of its facets. Instead, this research posits that surveillance is an act of structural violence situated and sited by its praxis. Within that praxis, the intentionality, materiality and purpose of surveillance should be considered in any act of

analysis or criticism, in order to understand the ongoing effects, and how the drive to surveill is continually recreated.

Definitions

At a fundamental level, surveillance indicates an act of watching. With an etymology from the French (Lyon, 2001:3) for 'watching over from above', the term originally had a degree of paternalistic care implied, and Staples (1997:ix) considers it simply as the '*act of keeping a close watch on people*'. However, this straightforward definition is not enough alone to classify the many differing approaches, technologies and systems of surveillance present in society. Instead, approaches to defining surveillance, or where on a surveillant spectrum a particular practice may lie, centre on what this research terms intentionality and materiality.[8]

Hier and Greenberg (2007:381) offer a definition of surveillance as "*the garnering and processes of personal information to regulate, control, manage and enable human individual and collective behaviour*". Here the intentionality is clear, in that 'processes' indicates a deliberate practice, while materiality is clear from the retention of the data gathered for re-use. Lyon (1994,2001) goes on to elaborate a similarly holistic view, stating that it is a "*focused, systematic and routine attention to personal details for the purpose of influence, management, protection of direction.*"

Materiality

Dandeker's view is primarily focused on materiality: He originally noted that

> *"[T]he contrast between capitalism and socialism, at least in respect of the administrative salience of bureaucracy, would seem to be one of degree. This is the context in which Max*

8

> *Weber developed his bleak view… as being enclosed in an administrative 'iron cage'"* (Dandeker 1990:2)

but later went on to say that surveillance was:

> *"(1) the collection and storage of information, presumed to be useful, about people or objects; (2) the supervision of the activities of people or objects through the issuing of instructions or the physical design of the natural and built environments; and (3) the application of information-gathering activities to the business of monitoring the behaviour of those under supervision and, in the case of subject populations, their compliance with instructions, or with non-subject populations, their compliance with agreements, or simply monitoring their behaviour from which, as in the control of disease, they may have expressed a wish to benefit..."* (Dandeker, 2006:225).

Whilst initially compelling, Dandeker conflates surveillant acts with those of other compulsion, bringing in an aspect of compliance that is not commonly seen in definitions of surveillance itself. Although compliance may be a desired outcome *from* a surveillant system, it is not a necessary requirement *of* a surveillant system. Dandeker's final expression of "*...behaviour from which… they may have expressed a wish to benefit...*" is unexpected here, and further reduces the delineation between monitoring and surveillance. In this single definition there are multiple and conflicting levels of compulsion, coercion and voluntary compliance with surveillance systems, which is unhelpful in terms of generating a typology or usable definition of surveillance.

Wall considers the issue solely in terms of materiality:

> *"Surveillance is "the act of monitoring the behaviour of another either in real-time using cameras, audio devices or key-stroke monitoring, or in chosen time by data mining records of internet transactions"* (2007:230).

However, this definition is too narrowly focused, considering only the act of watching - it makes no account of the purposes of such acts, or the purpose to which data intentionally gathered in such a fashion is used.

Intentionality

Surveillance studies is heavily influenced by Foucault's viewpoint that surveillance is an emanation and extension of the 'micro-disciplines', by means of 'coercion by observation' (Foucault, 1997:170). Foucauldian approaches rest primarily on intentionality through agency, positing that surveillance presupposes an intent to influence and control. These definitions tend to focus less on the 'micro-practice' of surveillance, the gathering and collection of data, and more on the broad motives and intent of the state actors. Placed in context with the previous section, they are more concerned with the intentionality and agency of the surveilling bodies, rather than how those factors came to be constructed and supported, or on how the interplay of levels of intentionality and agency affects that construction.

In terms of intentionality, Ball and Haggerty (2005:132-133) argue that surveillance should be taken to mean not just that someone is watching, but that there is a purpose, an additional intensity and intentionality not present with the casual observer, often with the assumption of negative rationales as to intentions.

For comparative purposes, Marx's definition is that surveillance is:

> *"scrutiny through the use of technical means to extract or create personal or group data, whether from individuals or contexts"*
> (Marx, 2005)

This has a useful resonance with the definition that has been provided by this research, but does not directly address the imbalance of power present between those carrying out surveillance and their subjects.

Surveillance has become so pervasive (Murakami and Webster, 2009; Rosenzweig, 2010) and will remain one of the key organizational structures of contemporary society (Haggerty and Ericson, 2000) that it may in some cases become unintentionally intentional. In practice, although there is no intent to exercise a form of structural violence against surveilled populations, the mechanisms of surveillance are simply the default way of arranging business, government and social structures.

This expands and builds upon Giddens' (1990) view that surveillance grew as a result of modernity. Giddens defends surveillance as a process in its own right, rather than simply as an emanation of capitalism, earlier arguing that:

> *"Surveillance as the mobilising of administrative power – through the storage and control of information – is the primary means of the concentration of authoritative resources involved in the formation of the nation-state"* (Giddens, 1985, p. 181).

For Giddens, surveillance is an inherent organisational feature of modernity and the modern state, one that is created and perpetuated through necessity of state administration. This reflects the intertwining of capitalism as an emergent and then dominant economic system with the liberal democracies that enabled and were facilitated by it.

Bogard offers that "*to surveil something essentially means to watch over or guard it. Guardianship is…an art of control that makes it safe for something to move freely"* (Bogard, in Lyon, 2006). Although he notes Lyon's view that surveillance both constrains and enables social relations, he goes on to note the emerging role of assemblages in considering surveillance. Considering a Deleuzean line of flight, he notes that: *"some lines of flight can become fixed in their direction, speed, intensity, eject… in the same way, deterritorialization does not always imply freedom.*" (Bogard, 2006: 101). The initial definition centres on paternalistic assumptions of normative good, that sit more comfortably with conceptions of monitoring and guardianship than with the stance taken on surveillance here. Adopting an assumption of normative

good for surveillant practice, that the objects under this art of control are ones that rightfully can, should and are controlled through surveillance, not only ignores the asymmetry of power in such relationships, but allows it to perpetuate by creating an assumption of 'rightful' asymmetry, explored in the following section.

Purpose

The pre-requisites of materiality and intentionality have already been proposed, and are validated by the literature. What also emerges, however, is a third prerequisite, that of purpose. Without materiality, intent and purpose, systems cannot justifiably be termed as surveillant. A derived characteristic of surveillance systems is the power asymmetry between surveilled and surveillor, which is present in some definitions of surveillance from existing literature, but not universally. This is relatively controversial in the context of surveillance studies, as it removes some surveillance-resisting activities and sousveillance from being surveillance. However, the very act of resistance by surveillance praxis should be taken instead as an 'anti-surveillance', creating a line of flight through existing systems. Activities typically identified as sousveillance attempt to invert or otherwise upset the power asymmetry of typical relationships. In doing so, they attempt to block or otherwise thwart the intent and purpose of surveillance systems, and cannot therefore be categorised as a subset of surveillance.

The ongoing contestation of surveillance terminology is highlighted in Ball and Haggerty as:

> *"merely labelling different sociotechnical relationships as 'surveillance' does little to enlighten us as to the dynamics of the control, resistance, emergence and development of surveillance practices."* (Ball and Haggerty, 2005:133)

Haggerty and Ericson (2000) advanced the concept of the surveillant assemblage: a multiplicity of surveillance systems and processes that are

becoming increasingly connected. This is despite – in many cases – the lack of a single or shared point of origin. Those systems, without any particular guiding power or authority, create 'data-doubles' of individuals. These are (to our eyes) poorly defined collections of personal data that have an existence separate to, and beyond, our own lifespan.

The concept of the assemblage has some considerable traction in surveillance studies, and the surveillant assemblage is a touchstone of the discipline, with over 600 citations. However, there is a significant area of work that can be carried out to advance this theoretical base, particularly in terms of power systems, public policy and hegemony: factors which have a strong bearing on the creation and dissemination of surveillance systems. Although there are common elements to a number of definitions of surveillance, there is still little in the way of consensus.

Surveillance as Inherently Negative

This research adopts a different approach. Having considered the main approaches to surveillance, the rest of this chapter sets out the rationale for using the 'negative' conceptualisation of surveillance outlined in the introduction, its value to critical surveillance studies, and how research can make use of that approach.

Fuchs (2011:109) argues compellingly that dialogue around the nature of surveillance is "*important to show commonalities and differences between various approaches.*" For Fuchs, the main determinant of surveillance is whether approaches conceptualise it neutrally, or whether it is approached as a praxis of domination and systemic violence. As already indicated, this research argues that current conceptions of surveillance are too broad, and in particular, that benign monitoring should be separated from surveillance, as lacking of the features of materiality from other surveillance systems.

For the purposes of this research, surveillance should be considered a confluence of the three prerequisites:

- **Intentionality:** the surveillance must be *deliberate*
- **Materiality:** the actions must gather *material information* about the subject, or *information that can be made material*
- **Purpose:** the intent of the actions is to enhance, allow, facilitate or exert the flow of power

As well as ruling out 'bottom-up' surveillance or sousveillance, this also rules out the bulk of, for example, patient monitoring – the classic example being caring for dementia patients (Kenner, 2008:252) or epidemiology (Hankey, *et al.*, 1999)[9]. Although patient monitoring may utilise similar mechanisms of surveillance, it arguably lacks the *purpose* that characterises a surveillant system.

The view of this example would change if, for example, even an ancillary purpose of the monitoring system was to create patient data banks that were used to inform insurance premiums. This would move the system across the boundary between medical or epidemiology management purposes (Tokars *et. al.*, 2004), and surveillance for actuarial purposes, which reinforces existing systems of capital and power.

Creating an Other

This distinction of surveillance and more benign monitoring is useful because it identifies the creation of the Other, the subject, body or target which is marked as not part of the majority group, or somehow outwith a broad sense of the normal within a particular group or context, while the concept of the Other is common in social sciences and political science, it also has resonances in contemporary and continental philosophy that is drawn on in this research.

Deleuze talked extensively about difference, considering that the centrality of difference to human thinking had been too long subordinated to the other

9

'pillars of reason', meaning that the subtleties of difference, and the networks of graded interactions that this created were overlooked in favour of Hegelian opposition (Deleuze, 1968). Derrida considered *différance* in terms of opposition, but rather than considering the opposition between the object and the wider world, considered it in terms of the *system* of differences, or "*the spacing by means of which elements are related to each other.*" (Derrida, 1981)

Later Deleuzian work saw a shift towards 'segmentarity', the societal arrangements by which spaces and practices are delineated - home and work, class, race and language coding. Deleuze and Guatarri (1987:245) go on to note that "*not only does the State exercise power over the segments it sustains or permits to survive, but it possesses, and imposes its own segmentarity."*

This differentiated self or body, which exists in a delineated space, is an essential (in its most literal sense) element of surveillance systems and assemblages: doubly so of surveillant assemblages. By attempting to understand what surveillance and creation of the Other might mean at particular levels of analysis, it is useful to consider the concept of social sorting, such as Gandy's Panoptic Sort (1993).

Surveillance systems collect or create data, which is then used to make decisions and selections about individuals, ranging from access to the welfare system, to credit decisions or loyalty discounts for shoppers (Whitaker, 1999). This process is inherently exclusionary: to have entitlement, there must necessarily be groups who are not entitled, which brings us to what is arguable modern surveillance's main use: the identification (and often, the creation) of the Other.

The composition and subjectivity of the Other governs human relationships, and therefore behavior at a fundamental level: "*in the beginning was the human relation.*" (Lacan, 1953 in Miller, 1993). Lévinas noted that *experiencing the other* constitutes both distance and proximity (Lévinas,

1987). The metaphorical or ideological distance of the other is experienced in their proximity to the subject. This means that although individuals may have a degree of intersubjectivity[10], they are still separate people, and the tension between these two aspects underpins the relationship between people. Surveillance, which records and categorises the characteristics of individuals, delineates the differences much more clearly. Systems of panoptic sorting have a potentiating effect, and actually impose difference on individuals by dint of that classification. In any system that classifies, for example, entitlement, the causal effect is to create an Other – those who are not entitled.

It is important to examine the distinction between notions of the 'not-other' as where 'same' could potentially be used. 'Same' indicates actual similarity, a number of shared characteristics, with an overlap and congruence – however, the 'not-other' notion indicates only the shared characteristic of simply not being characterised as other. This research wishes to make a tripartite distinction between the two elements of the separated and sorted population and *those administering the sorting* as three constituent elements, without which surveillant structures and social sorting could not exist.

In terms of the assemblage, which are explored more fully in Chapter Five, the Other is both an expressive and constitutive element (DeLanda, 2006). The surveilled subject's relationship with the other (and constructions of the Other, which includes behaviour modified by the presence of surveillant assemblages) creates both 'real' elements (changed behaviour) and expressive elements – the signs and signifiers used to convey the meaning of the surveillant system to the surveilled. The relationship between surveillor, surveilled and other is a crucial part of the multiplicitous nature of surveillance, and it is this which characterizes its presence as an assemblage. There is a significant parallel with Bourdieu's notion (1998) that the asymmetrical availability of societal resources is a strong differentiator of populations:

10

> *"The idea of difference… is … a set of distinction and coexisting positions… exterior to one another and defined in relation to one another through their mutual exteriority."* (ibid: 6)

That is to say: people are othered, sorted and managed according to their existing access to resources external to them (beyond simple economic resources), which in turn can dictate their ongoing access to further resources.

The creation, management and control of the Other is a key factor that identifies surveillance, and management of the Other runs through the praxis of intentionality, purpose and materiality already outlined. It is central to Wright's 'technologies of social control' viewpoint (2005), to Ball and Haggerty's issue with the labelling of certain socio-technical relationships as surveillance, and is the underpinning rationale for dataveillance (Clarke, 1988), the "*…systematic use of personal data systems in the investigation or monitoring of the actions or communications of one or more persons.*"

The Other and Categorical Suspicion

Contemporary surveillance is predicated on the assumption that anyone who enters a surveillance site's sphere of influence is assumed to be the Other until proven otherwise, an assumption that is heavily driven by political rhetoric (Frois, 2011), in a similar manner to Füredi's (2009) observations noted earlier on the prevalence of risk-management in society. Lyon (2001:333) notes:

> *"The problem here lies in the process described by Bauman as adiaphorization, a concept that first appears in Postmodern Ethics (1993) and continues use to the present (in Liquid Fear 2006, for example). This speaks to ways in which 'reasonable decisions' are declared morally indifferent, a process that begins in bureaucracy but is amplified in a world of self-augmenting*

technologies supported by the momentum of current political economies of panic, precaution, and outsourcing."

Such self augmenting technologies are predicated on, and further entrench, power asymmetries. These processes of *prima facie* reasonable decisions create a system of categorical suspicion, an inversion of typical judicial processes; rather than requiring the actor with the balance of power to show guilt or categorisation, the onus is often on the surveilled subject to show that they are 'not-other'.

Such 'categorical suspicion' (Marx, 1988) is increasingly easy to apply *en masse* as surveillance technologies become more advanced, more prevalent and more normalised (Lomell, 2004). The systems become easier to apply, even as the force wielded by the state and large techno-capitalist organisations becomes greater. This has two main implications in terms of power analysis. As categorical exclusion takes place, an increased number of places are spatially and systematically prohibited for the other. Power is exerted as default by those whose identity is securitised (Rose, 1999), who are also normatively those of the 'default' identity. This leaves those who are marginalised and excluded in the position of having to prove themselves to a system that is already built on an imbalance of power. This is an explicit and ongoing form of conflict, which DeLanda notes "...*has the effect of exaggerating the distinction between 'us' and 'them', that it is, it sharpens the boundaries between insiders and outsiders."* (DeLanda, 2006:58)

The other implication is that surveillance systems will track those who are subject to categorical suspicion, resulting in the ability to exercise power, that exists *in potentia,* at any time. The exercise of that power becomes a self-fulfilling prophecy, leading to the justification of surveillance, and creating a demand for further and more comprehensive surveillance. Given the dependence on the Other, as Lyon states:

"We need a sociology of surveillance for these "times of terror"' – a multidisciplinary approach that does not ignore how

> *indispensable surveillance is, but acknowledges both undesirable cons and problems in a just society."* (Lyon, 2010)

Here Lyon assumes that a society which regards surveillance as indispensable *can* be just, following the 'neutral critique' of surveillance noted by Fuchs. However, this research makes the argument that surveillance is one aspect of the whole, and it is the nature of the many-to-whole relationship that can and should be critiqued – meaning, *inter alia*, that a societal relationship predicated on the existence of surveillance, with its imbalances of power and structural exclusion, *cannot* be just.

In part, this is achieved by rejecting the notion that surveillance, as defined here, is indispensable. Rather, it is an emanation and territorializing symptom of the current political order; to successfully challenge surveillance is to successfully challenge the current political structures in place, and vice-versa. Although not indispensable to society as a whole, surveillance structures may well be indispensable to the current power structures, of which it forms a constitutive part.

Control and the Capitalist Other

> *"If we understand Foucault as saying that powerful actors control disciplinary power, then the notion of centralized and hierarchical surveillance is still valid. It is easier to exert counter-power, but there is an unequal distribution of power."* (Fuchs, 2011:119)

This highlights the similarities in the main difficulties faced in moving both surveillance studies and public policy towards more complexity based approaches. It is difficult to adequately describe the multiplicity of forces that are at work in any kind of 'centralized' or hierarchical system, much less systems that purport to be modern liberal democracies, by systematising inequalities of power. Instead, both disciplines need to understand the forces and processes by which assemblages are created, sustained and – rarely –

destroyed. In turn, this argues strenuously against the concept of a coherent centre that acts with any kind of consistent or unitary agency. This appears, at first glance, to be contradictory to the view held by Fuchs. However, it is important to understand both the assemblages that construct and constitute these powerful actors, as well as how those inequalities of power manifest themselves in the internal relationships of assemblages around the subjects of surveillance. The *appearance* of centralized and hierarchical surveillance, just as with government agencies and state actors, does not necessarily translate to the *praxis* of such centralisation or agency.

Surveillance, as laid out here, is a coercive act; it reflects a power asymmetry (economic or otherwise), and is never benign (Fuchs, 2011). Surveillance is an act of structural violence because of the systematised manner in which it reflects capitalist power structures, and segregates those of lesser use to the mechanic of those structures. DeLanda (2006:89) notes that "*systematic reliance on physical force… signals an unstable form of authority, so other material components must be added to these to align enforcement and legitimacy.*" Surveillance in late capitalist systems constitutes many of those material components. Although surveillance can, and does, lead to acts of physical force emanating from assemblages of the state to those surveilled, in the main, it represents a shift from physical manifestations of structural violence towards more implicit processes maintaining those same state-level assemblages.

Within conceptualisations of surveillance that are consistent within the definition laid out in this thesis, the asymmetry of power between surveillor and surveilled is apparent. It is a systematic emanation of the capitalist state assemblage, one which both constitutes and represents it. From a strictly utilitarian point of view, it would be interesting to assess the 'benefits' of surveillance against the costs and damage to agency, liberty and privacy that it brings. Because surveillance is intrinsically linked to mechanisms of structural control, it represents a form of structural violence, another "*avoidable impairment of fundamental human needs.*" (Galtung, 1969) Most

benign or ‘positive aspects’ of surveillance should instead be gathered under monitoring.

Making such a distinction has three main implications. The first is a degree of clarity around normative issues. Surveillance, as defined here, is centred around power asymmetries, characterised by the attributes of intentionality, materiality and purpose. These necessitate a particular normative approach to surveillance practice and critique, which addresses the complexity of the relationships between those attributes, and between the components of the assemblages that manifest surveillance systems. The second is that, in understanding surveillance as simultaneously a symptom and driver of late modernity, critique can be focused on deterritorialising negative assemblages, and it is here that such analysis begins to see significant movement in understanding ‘good’ policy. In short, the process of understanding and deconstructing assemblages that manifest in structural violence or domination can be a useful addition to the public policy process. Lastly, by removing the ambiguity of monitoring or positive surveillance from consideration, this allows the isolation of the issues and practices that occur at the junction of individual and state assemblages, which has great critical potential.

Chapter 5:

The Coalescent State – Policy as an Assemblage

Introduction

Assemblages are multiplicities which make up a whole. They are an organismic metaphor for systems and relations, where the overall nature of the system is characterised by the relationship between its heterogeneous parts (Dosse, 2012:138), and defined by relations of exteriority. If a part is removed from the system and placed in another, the relations both between the part and the new system, as well as the remaining parts of the existing system, will be altered, which makes them *intensive processes.*

From Multiplicities to Assemblages

In a philosophical context, multiplicities are complex systems that do not flow from a prior unity or existence. Instead they are constituted by the totality, or sum of the components, whether terms of numbers or consciousnesses (Bergson, 1910).

Deleuze felt it important to differentiate between the indefinite and definite – or absolute – multiplicity. While there can be an indefinite multiplicity, a multiplicity of a thing, he considered that the absolute multiplicity – 'a' multiplicity, a noun rather than a verb to indicate the progression of thinking from the 'one and multiple', which has a degree of essentialism contained within it. Instead, the absolute multiplicity is a different epistemological route of thinking, not predicated on the progression of a dialectic process to discover a single essential 'truth'. Even within a dialectic process, that

Delueze refers to as the ‘one to many’ movement, there is no duality, but instead, in his work examining Foucualt, Deleuze sees a series of multiplicities that structure the space of possibilities relating to the assemblage in question (Deleuze, 1988:34-36) .

It is important to elucidate the differences between assemblages and multiplicities. An assemblage is a particular arrangement or interaction in a multiplicity, characterised not just by the attributes of the multiplicity, but instead:

> *"...(a) multiplicity which is made up of many heterogeneous terms and which establish liaisons, relations between them”* (Deleuze and Parnet, 1987:69)

For the purposes of this work, it is necessary to provide a more concretised version of the phenomenon or concept than is perhaps readily accessible from *A Thousand Plateaus.* Bogard provides:

> *“Assemblages are not discrete objects, but consist of open relations among heterogeneous elements whose only unity derives from the fact that they operate together (and even then it is not exactly right to speak of a unity, since counteractualisations (sic) of the assemblage are internal to it and continuous). Every assemblage is a multiplicity composed of other assemblages that are also multiplicities that together form a functional, every-changing ensemble.”* (Bogard, 2006)

In laying out the relational nature of assemblages, Bogard is also quick to note the self-referential and recursive nature of both assemblages and multiplicities. This focus is central to considering assemblages in any context, but particularly in surveillance studies or policy analysis. By adopting this approach, criticism can be focused on the processes of transformation and change, rather than essentialist approaches that inevitably capture the attributes and character of structures of power.

These definitions and the subsequent impact of such a structure can be stated as:

> *“Assemblages are systems of 'things' that interact with each other. These interactions produce emergent effects: effects that are not easily analysed as the sum of their individual relationships or effects.”* (Campbell and Van Brakel, 2015)

Assemblages are intended as counter-structures and contrasts to traditional ‘arboreal’ epistemologies. These traditional epistemologies tend towards top-down, binaristic modes of operation, and are characterised by the attributes of the structure – the structure is reified as the object and considered holistically, rather than unpacking relations that result in an emergent characteristic.

These ‘new’ assemblages are instead rhizomatic or weed-like (Deleuze and Guattari, 1987), drawing relationships between themselves in relatively unstructured paths and patterns; they are a particular kind of relationship that develops “*‘au milieu’: in the middle, in between*” (Holland, 1991:98).

DeLanda offers that assemblages are ‘*the main theoretical alternative to organic totalities*.’ (DeLanda, 2006:10). These assemblages are wholes that are characterised by their ‘relations of exteriority’, meaning that an assemblage is a collection of bodies and components that have a rhizomatic relationship to each other, within a given context. An assemblage is both an object and a process. Being rhizomatic, each component (or 'assemblant') has, and can, create relationships with every other component, as well as to the whole:

> *"In assemblages you, you (sic) find states of things, bodies, various combinations of bodies, hodgepodges; but you also find utterances, modes of expression, and whole regimes of signs."* (Deleuze**,** 2007, pp 176 – 179)

These components can be removed from an assemblage and associated with another, but will create different relationships, both between the component and the whole, and between the whole and its other relationships. Crucially, these '*relations do not have as their causes the properties of the [component parts] between which they are established…*' (Deleuze, 1991).

Malins notes that the work of Deleuze and Guattari "*is perhaps best conceived of as a 'tool box'… a collection of machinic concepts that can be plugged into other machines or concepts and made to work.*" (Malins, 2004:84). He goes on to note that, through using these tools, "[*w]hat matters is no longer the subject or meaning of the bodily assemblage… but the specific affects it enables.*" (*ibid*.:94).

For the analysis of the specific subjectivity of the intersections of public policy and lived surveillance, this means that the internal rationality that produced the public policy is deprecated, in favour of understanding the affective output of the system that it has created or input to. Similarly, the systems of surveillance themselves have their internal rationality pushed aside, in favour of understanding how they have performed, not against their own metrics, but in terms of the impacts, experience and assemblages that they output. The outcomes and lines of flight generated by these assemblages are, in this analysis, more important than any of the inputs or constitutive processes. DeLanda perhaps states this most succinctly:

> *"The subject or person emerging from the assembly of subpersonal components… has the right capacities to act pragmatically, as well as socially, to select ends for a variety of habitual or customary reasons that need not involve any conscious decision."* (*ibid*:52)

On the face of it, this seems to contradict the position reached when considering policy: that decisions are not necessarily rational, particularly those taken by actors who have a system imposed on them through policy. The key phrase here is '...*emerging from the assembly of subpersonal*

components...': Individuals are constituted of their own assemblages, which then interact with larger assemblages. These interactions may appear, and in fact *may be* rational in their particular instance, but the congruence of assemblant factors should not be confused with the application of a unitary and common understanding of rationality.

Arguably, following Srnicek and Malins, the application of assemblages, as originally discussed by Deleuze and Guattari around texts, has potential that extends far into policymaking and statecraft; by extension, this application asks similar, and significantly valuable critical questions of surveillance and surveillant assemblages.

By varying the weight attached to the various levels of assemblages, and their interrelations, laid out in this paper, this research argues that it is possible to build a realistic and pragmatic understanding of the policy process through assemblage theory: policy as an assemblage that produces real material effects, ratter than solely transmitting information:

> *"An assemblage, in its multiplicity, necessarily acts on semiotic flows, material flows, and social flows simultaneously (independently of any recapitulation that may be made of it in a scientific or theoretical corpus). There is no longer a tripartite division between a field of reality (the world) and a field of representation (the book) and a field of subjectivity (the author). Rather, an assemblage establishes connections between certain multiplicities drawn from each of these orders... In short, we think that one cannot write sufficiently in the name of an outside."* (Deleuze and Guattari, 1987:27)

The assemblage, therefore, expresses itself into the world ('reality') in its expressive characteristics, and exerts an impact both on the *thing that represents it*, and that which is its subject; for policy considerations, and surveillant assemblages, this relation of the expression into the real world and to a subject is repeated near infinitely, with every subject of the policy.

In terms of utilising this to provide critical insight, Srnicek argues:

> *"this model of reality can be abstractly analyzed into three realms: (1) the actual which consists of the stable, identifiable systems and individuals that tend to cover over (2) the intensive process of individuation that produced them, consisting of 'far-from-equilibrium' processes that are 'metastable' and that embody (3) the virtual structure of potentialities that are immanent to a situation…. While we can break them apart for convenience, each is real and always in a concrete mixture with the others."* (Srnicek, 2007)

This supports the approach taken here – the identifiable systems of government and policy often overshadow and block the analysis of the *actual* processes that go into creating them, as they are fundamental assumptions and patterns of existing and being. Around those patterns are the assemblants of the wider system, with many assemblants persisting and being present in many systems; the challenge of policy change, of materialising a *new* emergent system, is effectively impossible with acknowledging, and in some instances *changing*, the field of assemblants (or potentialities, as here) that is immanent to the situation.

The surveillant assemblage advanced by Haggerty and Ericson (2000), and widely accepted within Surveillance Studies, has much to support it. It understands that, even now, no single surveillance tool has 'won out', and indeed there are a multiplicity of tools and practices. These overlapping regimes of surveillance have resulted in a *'a rhizomatic levelling of the hierarchy of surveillance, such that groups which were previously exempt from routine surveillance are now increasingly being monitored'* (*ibid*.: 606). In addition, the subjects of surveillance are first reduced to data, and then reconstituted as data doubles, informatic doppelgangers that can lead quite different – and troubling – lives.

Terminology

Assemblage theory has its own discrete, and at times dense, terminology. It is appropriate to set out the particular selection – and extension – of terminology used here, particularly in those areas where terminology and definitions are conflicted, or overlap with scholarship outwith this area. Here *coalescence* is used to describe the 'natural' state of the assemblage, as it is made and remade by its assemblants, but without pressure or strain. Coalescence is the drawing together and operation of an assemblage, on an ongoing basis. It becomes crystallisation at the point that an assemblage has a material impact, or generates its emergent effects: within the policy space when a decision is taken or avoided. The point of crystallisation is the point at which the policy assemblage territorialises itself over its subjective elements and exerts 'control', either by enacting a policy, or by excluding some of the subjective assemblants from power structures.

Territorialisation and deterritorialisation are key concepts. Broadly, territorialisation is the set of processes and relations that make an assemblage or rhizome more stable (Campbell and van Brakel, 2015). Bogard's (2006) description of territorialisation: "*staking out an 'event-space' where flows are made to pass in a stable fashion",* could be tailor-made for the purposes of understanding policy as an assemblage. Protevi (2012:254-256) holds that territorialisation affects the rhythms of the assemblant components (effectively, the way and manner in which assemblants are translated from one assemblage to another, a concept which is not used or developed further here). Through the act of territorialisation, these assemblants (and as a consequence, the overall assemblage), have both dimension and expressiveness – an impact and a 'value', that can be elucidated, enumerated, or otherwise circumscribed by language. This has the effect of intensifying the assemblage, which is returned to as the concept of crystallisation.

Finally, it should be noted that this research uses *territorialisation,* rather than the more strictly Deleuzean *'reterritorialisation'.* Deleuze used reterritorialisation in earlier works to differentiate it from the Lacanian concept of territorialisation as relating to the personal and libidinal body (Holland, 1991). In later use, reterritorialisation gives way and is used somewhat interchangeably with territorialisation, as it retains its concept of being the '*dead hand of the past'* (ibid:62), while also being imbued with additional meaning, around the state being one of the most significant locations of territorialisation. This is particularly the case around colonialism (and the collapse thereof), in relation to both the physical spaces and emergent systems that they imbue those spaces with.

Emergence is the final key concept that must be laid out. Significant reference is made to emergence and emergent systems in both Deleuze and Guattari's work. Protevi holds that it is the construction of a specific functional structure (order) at a point in time, from a series of successive points in time; it is the specific instantiation of the particular system under discussion at that moment in time (Protevi, 2006:19). Emergence is important as it allows the *specificity* of an assemblage to be discussed and analysed, both in terms of the manner in which that assemblage was constructed and created from the complexity of its multiplicitous relations, and as a unitary whole in its own right.

There are some parallels with Latour's (2007) Actor Network Theory (ANT), which although not examined in depth here is a useful parallel to the theoretical themes developed here. Moving the ANT discussion into the postmodern terminology is useful for several reasons. In the first instance, the concept of policy as assemblage does attempt to understand how those assemblages came to be (and to a degree, the 'why', although this may not be an apposite question), which ANT tends to avoid. Secondly, Latour himself noted the similarity, but noted the clumsiness of talking about "actant-rhizome ontology". In the last instance, the criticisms of ANT as overly descriptive, whilst overblown, have some resonance here. Bringing the concepts of

fluidity, movement, coalescence and crystallisation to the assemblage attempts to deliver a repeatable method that provides critical insight, and which is relatively unburdened by previous epistemological debates within the field.

An assemblage, as a collection of heterogeneous assemblants, can be presented as a *circumstance* or *context* within which those assemblants exist. They are maintained, changed, and in some circumstances, created by (social) processes of territorialisation and stratification. Assemblants can be part of many assemblages, while assemblages can (and usually are) themselves be part of other assemblages: each assemblage is itself *recursive*, in that it is continually made up of its parts. It is the substantive relationship between the parts and the whole that make up the character, the context, of the assemblage. In the strictest sense, assemblages are never *created*. Instead, they grow from smaller assemblages, the collocation and collaboration of a number of recurrent processes of production.

The rhizomatic nature of surveillance - that it grows like weeds - is central in moving from the concept of policy as an assemblage to the subjectivity of surveillance. However, the conceptualisation of *policy and environmental context – the assemblage* that allowed this growth to take place – has not yet been adequately theorised. Instead, the focus has been on the output of surveillant assemblages, rather than the process and structures by which those assemblages have been created. Haggerty and Ericson's previous study (1999) into the dispersion of military technologies into civil spheres can, in this light, be recast as a deterritorialisation of those civilian spheres, and a crystallisation around the new technologies.

Surveillant assemblages provide us with the link into policy that can operationalise such a theoretical framework. The surveillance sphere has the opportunity to examine comparatively short causal chains between policy 'decisions' and subjective experience. The experiences of protestors and activists in the UK in recent years, particularly those affiliated to UK Uncut and

Occupy London should allow us to examine the near term impact of decisions and gain critical insight into the operation of the state as an assemblage.

The Coalescent State: Policy as Assemblage

Policy has all of the characteristics of an emergent system. It is complex, with properties (or emanations) that are not directly reducible to the components of that system. Accordingly, essentialist approaches to policy should be rejected, particularly essentialist-rationalist conceptualisations. Policy approaches, to date, tend to reify attributes of policy. Srnicek notes that *"...the individual-system takes on a stability that lends it a sense of solidity, and permits theorists to draw out its 'essential' properties, without which the system would become something different.*" (Srnicek, 2007:43)

Returning to the definition of government as "any more or less calculated and rational activity" (Foucault, 1991:29), the key issue is the 'various forms of thought' which accompany it. Arguably, any activity that is 'more or less' calculated will always tend towards the lesser, towards entropy: Government tends towards a diffusion of agency and, generally, of intent.

> *"…the way forward for critical social policy is to reconfigure governmentality and adopt a 'realist perspective".* (Stenson, 2005)

Rationality, or even agency, cannot necessarily be directly ascribed to government, because they are made up of too many divergent and moving parts: the stability and consistency of the act of governing is far from certain. There are a multiplicity of authorities at play both within and outside government that have impacts and influences on policy, through the interplay of their assemblages with those of government and the state.

The multiplicity of authorities constitutes an important part of the issues that theory seeks to unpick: how the techniques and epistemology of state varies from context to context, and from level to level. Acknowledging this explicitly

introduces a large degree of complexity and recursivity, which needs addressed. In turn, we choose to embrace that make it part of the overall toolkit:

> *"…political science has largely remained bound to ontologies which privilege simple and static entities. Their very suppositions about the nature of reality tend to reflect a previous time in which clarity and simplicity could (more plausibly) be considered intrinsic properties of the world. Most glaringly, rational choice theory often presents itself as 'a new master social science' capable of a single, comprehensive analysis uniting political, sociological and economic behaviour. Its reliance, by its own admission, on axiomatic, unnaturally perfect conditions make it a frighteningly poor tool to analyse the complexity of contemporary politics."* (Srnicek, 2007)

If governmentality is "...the ensemble formed by the institutions, procedures, analyses and reflections" (Foucault, 1991:26), then an obvious criticism is that such a framework is not significantly different to that ensemble. However, assemblage analysis combines two main aspects - temporality and recursivity. Assemblages coalesce, crystallise and interact over time. Each assemblage is the product of a previous assemblage. Each component is part of many assemblages. The relationship between the assemblant and the whole it resides within is the unit that provides the analytical basis. This is a realist/pragmatist focused approach that allows both theoretical insight, and lends itself to a mixed methods approach to empirical work that moves beyond the simple to provide empirical support for those critical viewpoints.

> *"…this mixed methods approach gives more attention to the empirical concerns of social policy by examining particular mentalities of rule in their local context. In doing so, it renders visible the actual effects of governing practices and the behaviour and situated knowledge of subjugated populations."* (McKee, 2009)

"*..the analysis of micro-powers, or of procedures of governmentality, is not confined by definition to a precise domain determined by a sector of the scale, but should be considered simply as a point of view, a method of decipherment which may be valid for the whole scale, whatever its size.*"
(Foucault, 2008:186)

Policy as Assemblage

Policy is both a 'thing' and an on-going process. Like many multiplicities, it is characterised by the 'part-to-whole' relationship. The policy system has processes that, while not essentialist, are intensive in a Deleuzean sense: they cannot be changed without altering the emergent characteristics of the policy assemblage.

"Processes exhibiting intensive properties are those that (1) cannot be changed beyond critical thresholds (the 'line of flight') in control parameters without a change of kind (a 'becoming'), and that (2) show the capacity for meshing into 'consistencies', that is, networks of bodies that preserve the heterogeneity of the members even while enabling systematic emergent behaviour."
(Bonta and Protevi, 2004:15)

This research suggests that policy is the whole which emerges from its parts. The policy process is the recognisable manifestation of the varied, unequal and often asymmetric relationships between its assemblants; arguably, it is an archetype of macro-scale assemblages. "Society" would be too broad a stroke to paint - although it is an assemblage, it is assemblage layered upon assemblages, the epitome of an '*entity under study… composed of parts operating at different spatial scales*' (DeLanda, 2006:32).

The issue, both for policy and society (and indeed, the specific part-to-whole relationship that characterises their interactions), is understanding the massively recursive nature of these scales – as policy emerges and comes

into existence from the interactions amongst its assemblants, it begins to exert actions against those parts, changing the relationship and beginning a new recursion. Eventually, these policy systems become meta-stable – they are subject to little or no change *at the upper macro levels of analysis*, with only minor levels of changing relationship occurring at lower levels. At this point, the system has become territorialised, and the lines of flight and other deterritorialising forces that work against that system's cohesion are minimal, or occur at a low-enough level that they do not endanger the 'whole' system – in our specific instance, a policy.

Coalescence and Crystallisation: Decisions and Non-decision

Public policy tends to highlight 'decision points' as key demarcations, and although these would also include points where decisions have been explicitly or implicitly *avoided*, that terminology is used here to encompass all of these variants:

> *"Policy choices are frequently made in the absence of a clear-cut, once-for-all decision. They simply 'happen', in the sense that certain steps are taken that are necessary…"*
> (Bachrach and Baratz, 1970)

This is the point at which the assemblage around the policy *crystallises*: the practical effect of the assemblage at a given point in time, it has some kind of 'output', and in many cases will create subjective experience, analogous to the 'Moment of Exposure' (Ball, 2009). The process of crystallisation generates the emergent effect of a policy assemblage, and feeds into a new policy assemblage. In the case of government, this has many of the same assemblants in place, highlighting the broad meta-stability of such assemblages. This is differentiated from Policy Network Analysis by temporality and recursivity. As outlined above, policy networks are not static, and recur, reform and persist in *different forms and configurations* across time.

> *"Instead the concrete actualization of events results from the interaction of diverse causal tendencies and counter-tendencies. Now, whilst it may be tempting to argue that this interaction itself can serve as the single causal mechanism that necessarily generates the necessary happening… such interactions cannot be attributed to the operation of any single causal mechanism. For these to result from interaction among diverse causal tendencies and counter-tendencies. This opens the route to an infinite explanatory regress into the path-dependent past."* (Jessop, 2005)

The problematisation of 'infinite explanatory regress' can be rejected, not out of hand, but on two bases. The first is that, as mentioned above, some assemblages, particularly those of government and policy can be regarded as metastable; recursively, this limits both the *value* of infinitely regressing through time and *the necessity* of doing so. Where the assemblage values or outputs have changed allows us to trace the entry or impact of lines of flight. It also allows a reduction in the number of recursions needed to examine to determine our critical interventions and units of analysis.

Government, in this context, can be viewed as an assemblage that is metastable. Actions, at an overall meta-level can be predicted with a reasonable degree of certainty and comfort. That certainty feeds into policy assemblages in the form of actors, institutions, viewpoints and existing public discourse. There is a significant degree of limiting and confining assemblages in place around systems of government in the UK (and the West more generally).

Although the system of Government is nominally 'open', as Srnicek suggests, the practical reality is that the '*force of equilibrium*' – or what can be termed *'patterns of behaviour that emerge within a system*' – are extremely difficult to overcome. In Foucauldian terms, the system of power is very closed. It is highly territorialised through a monopoly, not just on state violence through the police and security services, but through literal access, and some degree

of control over the law-making process itself. Policy-making and particularly policy implementation then become more of a process of adjustment within the confines of this heavily territorialised system. The legislative and policy assemblage must move fast enough to '*capture the agreements while they last*' (DeLanda, 2006:43). Although crystallisation is the fixing and territorialisation of an assemblage for a given period of time, understanding that temporality, and how it adds or detracts from the territorialisation of *higher level assemblages* is crucial in understanding policy from that point on. This has a number of implications for surveillance studies.

The first is that the surveillance state as theorised and popularly imagined *does not exist.* The *de facto* reality for many may not be significantly different, but the process that creates that reality is very different. What are generally perceived as 'surveillance states' or 'surveillance societies' are instead confluences and crystallisations of overlapping self-interest and conflicting rationalities that play out in different manners, but have one main effect: the distribution and infinite dissemination of surveillance technologies and capabilities.

Kullenberg and Palmas note DeLanda's early thoughts about the availability of 'panspectric' technologies:

> *"...instead of positioning some human bodies around a central sensor, a multiplicity of sensors is deployed around all bodies… [It] does not merely select certain bodies and certain (visual) data about them. Rather, it compiles information about all at the same time, using computers to select the segments of data relevant to its surveillance tasks."* (DeLanda, 1991:206 in Kullenberg and Palmas, 2009)

However, this sits uneasily with his later work, as cited here; rather than viewing these panspectric technologies as assemblant components, they are viewed as Kullenberg and Palmas posit:

> *"[T]he panspectric diagram may be understood… along the lines of its effectuation in concrete assemblages. Both diagrams consist of paradigmatic sets of technologies, architectures and material components."* (*Op cit*.:3)

These effectuations must be thought of in the more holistic policy-as-assemblage analysis context, as one of three things: Assemblants in their own right, which exert relationships against the whole, and consequently on other assemblants; accelerants that affect the part-to-whole relationships of other assemblants, while having no existence within the assemblage themselves; or coalescing technologies that territorialise their own assemblages of data and meaning around them.

It is this last possibility which is most compelling. Not only do the panspectric technologies represent a territorialising assemblage in their own right, affecting the assemblant parts they surveil, but their territorialised assemblages form part of the larger surveillance structures that surround us, and make up the panspectric technology, as envisaged by Kullenberg and Palmas.

The second is that meaningful entry points to the policy process (and by extension, to contestation of that protest) are multiple, contested and often hidden: the policy process as previously conceptualised has not come to bear on the decisions that have been made on surveillance capabilities and targeting. In this context, understanding the 'attractors' – or forces for equilibrium – is vitally important. Again, these are not characterised by the actors or institutions, but *by the relationships of those actors and institutions to others*.

Cracking the Coalescent State: Methodological Implications

> *"Follow the plants: you start by delimiting a first line consisting of convergence around successive singularities; then you see whether inside that line new circles of convergence establish*

> *themselves, with new points located outside the limits and in other directions. Write, form a rhizome, increase your territory by deterritorialisation, extend the line of flight to the point where it becomes an abstract machine covering entire plane of consistence."* (Deleuze and Guattari, 1987:11)

The concept of policy as assemblage is a compelling one, in part because it is far from 'neat'. It acknowledges and incorporates the chaotic and at times inchoate nature of policy-making. It may be that, to make best use of such a theoretical framework, some kind of methodological approach that similarly builds on empiricism the fundamental complexity of policy is necessary. International Relations utilises the 'process tracing' methodology (Lapid and Kratochwil, 1996), and it may be possible to apply a similar process to the question at hand:

> *"The process-tracing method attempts to identify the intervening causal process – the causal chain and causal mechanism – between an independent variable (or variables) and the outcome of the dependent variable … Process tracing forces the investigator to take equifinality into account, that is, to consider the alternative paths through which the outcome could have occurred, and* ***it offers the possibility of mapping out one or more potential causal paths that are consistent with the outcome and the process-tracing evidence in a single case."*** (Bennett and George 2005:206-07, emphasis added.)

This is not to provide a typology or taxonomy of assemblages, but rather a conceptual framework – and perhaps some practical methodologies – of understanding what assemblant relationships within assemblages of policy can tell us, both about surveillance policy, and more widely. *"We must search for the genesis of a particular given itself."* (Srnicek, 2007:37)

> *"Methodologically, this requires a 'method of articulation' that respects contingent necessity and complexity. One way to*

> *understand this is to see it as based on the dual movement from abstract to concrete along one plane of analysis and from simple to complex as more analytical planes are introduced in order to produce increasingly adequate explanations"* (Jessop 1982:213-19)

The assemblage exists (in fluid form) between two points in time. The first is at a decision point where one crystallisation led to the existing assemblage. The second point in time is when the transformation of the current assemblage is complete: it becomes and coalesces into another assemblage, and so on in recursive fashion.

Identifying these decision points allows us to gather data on the various assemblants in place at any given time. These assemblants can be termed the 'coalescent potential'. We would theorise that, as an assemblage gets closer to a decision point, the assemblants 'react' (in a dialectic manner, although it is multi-dimensional rather than bilateral), to produce the dominant relationships that determine whether the assemblage persists the status quo (territorialises it) or acts against it (deterritorialises it). These dialectic processes also determine whether or not lines of flight are created, and through deterritorialisation of the assemblage, change the emergent effects and characteristics of the assemblage. In turn, these changes are visible in subsequent assemblages.

This lays out a framework within which it is possible, and indeed desirable, to theorise the complexity of the policy process and policy analysis. This raises implications for methodological work, since approaching complexity 'head-on' requires some new and potentially profound alterations to methods.

Such complexity gives rise to new challenges both in terms of *analysing* the material (including assigning, where appropriate, values to information), and also in *criticising and presenting* the material. While DeLanda notes that

> *"Assemblages are characterised along two dimensions: along the first dimension are specified the variable roles which component parts may play, from a purely material role to a purely expressive one, as well as mixtures of the two. A second dimension characterises processes in which these components are involved: processes which stabilise or destabilise the identity of the assemblage (territorialization and deterritorialisation)."*
> (DeLanda, 2006:18-19)

Srnicek (2007), building on this, argues assemblages have four axes, with DeLanda's original position being expanded to take account of the scale of the population. Given DeLanda's repeated focus on and analysis of the many-to-whole relationship, this omission in the original work is a strange oversight.

- Types of roles – expressive or material
- Types of processes – territorialising or deterritorialising
- Degree to which expressive elements code or decode identity
- Distinction between 'molar' and molecular populations

However, there is a strong argument to be made that these axes can be reduced in number. The molar/molecular distinction, which is effectively a question of scale of the assemblage in question, is arguably an intensive property of the assemblants at the time of crystallisation. Similarly, the types of processes are defined by the relations of exteriority between any given two (or more) assemblants. Taking the view of assemblages as oscillating between points of crystallisation the process type is characterised by the line of flight or territorialisation from the previous assemblage.

This strongly implies that critical analysis of this nature requires multi-dimensional thinking. However, in practice, in spite of the infinite number of assemblages that could be created, only a limited range are visible or evident. It is pertinent to ask why.

One reason is that although assemblages are drawn from the totality of what is available; *they are not drawn at random*. As indicated in the previous discussion on the concept of process types within assemblages, current assemblages are affected by decisions and crystallisations of previous assemblages, in some cases from many decades ago. In attempting to find methods to support this theoretical framework, the following are relevant questions: *Why do some assemblages keep their shape and others do not?* Clearly, some play of territorialisation and deterritorialisation is relevant, but what makes some assemblages *successful* in territorialisation, where others fail and are deterritorialised into annihilation?

Although an effectively infinite number of assemblages can be theorised, only some are visible, and these constitute part of what can be termed the 'totality of what is available'. To understand this, some concept of inheritance and persistence is required.

Assemblages inherit components, and relationship sets from previous assemblages, whether by dint of accretion through territorialisation, or by the shattering of some previously understood paradigm by a line of flight. It is posited that there is a degree of commonality in the assemblants that are shared from one assemblage to another; a degree of overlap, shared vector, or some other similarity that allows the reshaping of assemblages – like policies – in similar fashions over time.

At some point, this can be regarded as a form of inheritance. Some combinations of assemblants (whether material or expressive) are so well embedded that they form their own assemblage, which interacts with successive larger assemblages over time, becoming a form of collective memory, a mechanism for persistence. Not only does this go some way to explaining why assemblages do not deterritorialise completely, or form in new and random fashions, but it also allows us to avoid the question of complete epistemic relativism. *Some forms and assemblages of knowledge* are so pervasive and persistent that they form the basis of *de facto* prior knowledge. In a loose allegory, the standard model of physics has been strengthened

through the discovery of quarks, tau neutrinos and the Higgs Boson. Each of these discoveries had the potential to completely deterritorialise the previous assemblage of knowledge, as a revolutionary line of flight, had they been found to operate in an unexpected way. Until a discovery is made that *does* constitute such a line of flight, then this knowledge can be regarded as *effectively objective*: even if one were to adopt a relativist position on all knowledge, there is little epistemic value to doing so.

How could these concepts of inheritance and persistence be measured, if they represent, in some cases, the boundaries of what is known and understood? In the first instance, consider what constitutes an assemblant. This can be held up as constituting elements and values. An element is a value-free component of the system, while the value is that characteristic expressed *in relation to the rest of the system.*

A trite example is the statement “I am very unhappy”. The statement expresses a simple sentiment in English, is easily understood, and is an expressive element. However, in the context of a wider (but still discursive) assemblage where other students are enthusiastically and happily discussing a holiday trip, it represents a strongly negative assemblant – one that may change the whole system, depending on the speaker’s worth and influence in the group. These additional factors are also assemblants in this system, albeit ones that are more difficult to circumscribe for the purposes of analysis, but the statement is set *against* the other assemblants of the group, and strongly so.

That same statement, observed in a separate assemblage of people making complaints about government cuts would be *with* the other (or at least, most other) assemblants, and may in fact be quite mildly stated in comparison to others.

Taken in itself, through these small-scale examples, this is not a revolutionary concept. The linguistic concept that context determines meaning has long been present. However, this type of analysis is almost scale agnostic,

because it focuses on the relationship between the individual assemblant and the overall system; some assemblants are people or processes, while others are organisations. Some individuals operate only in 'low-level' assemblages, which feed into other assemblages through officials or others, which also provides a useful model for mediated experience through the lens of assemblages.

Fundamentally, the rationale for this approach is that assemblages, in their widest senses offer significant critical insight from understanding the relationship between the immediate components and the system. The components are, effectively, vectors with a constituent element, which expresses both a magnitude and direction to the system in its own right. *What* these assemblants have a magnitude of depends on their nature as expressive or material elements.

Tensions and Uncertainties

As laid out here, the theoretical framework makes a firm delineation between forms of surveillance that should be the target of critical focus, and those more benign forms of monitoring that fall outside of this definition. Combining this sharply-edged definition with an overall theoretical approach that is reflexive and constantly self-constructed is challenging. On the one hand, such a categorisation seems to apply a very definite attribute to a system of relationships that is held here to be the product of those relationships. The theoretical framework laid out here indicates that the effects of such assemblages can only be seen in their operation, by the continual performance of the relationships within them. The assemblage is defined by the relationships of exteriority between its assemblants; in many respects, it seems that the application of a single, knowable classification is a reification of a particular attribute, which is embodied over the totality of the assemblage.

This is a tension that may remain using this type of theoretical approach, and remains as a caution to the practitioner, particularly in terms of assigning

causality between assemblages over time. As the assemblage is a multi-dimensional consideration, across expressive and material axes, taking account of these over time is necessarily complex, and could easily lead to incorrect assignations of causality: that some lived experience has resulted from the particular formation of an assemblage at a particular time, after failing to account for some particular aspect.

However, there are several factors which can be taken to mitigate against these risks or uncertainties. These can be broadly categorised in terms of *emergence, intensity, and metastability.*

The concept of emergence must be centred in any analysis. Focusing on the effects that *"are not easily analysed as the sum of their individual relationships*" (Campbell and Van Brakel, 2015) is both necessary and challenging. The construction of assemblages as counter-structures to traditional epistemologies requires the structures and processes that generate assemblages to be considered organically, with constant unpacking of those relations that result in the emergence of the effects and influences that researchers may wish to study.

The notion of intensitivity in Deleuzian thought brings with it issues of indivisibility: the components of the assemblage cannot be changed without changing the nature of the emergent properties. In the case of surveillance, a change to an assemblant within that surveillant system, for example a technology or practice, cannot be altered without changing the system. The focus of this, for the researcher, should be to examine the territorialisation of power asymmetries, and whether these differentials have increased or decreased.

This seeming conflict becomes particularly apparent on the borders between surveillance and other forms of monitoring. Given the multiplicity of relations that govern the intent, materiality and purpose of any potentially surveillant structure, some phenomena will necessarily be sited on the boundary between the two categorisations. In many cases, they will fall into both

categories at different times, which is both a strength and a weakness. This aspect is frustrating for the researcher as the lack of solidity, or seeming inability to satisfy a categorisation means that the permanence of a classification is impermanent at best. However, this same impermanence offers the opportunity to study the changes in relations, and therefore of emergent effects, that cause the structure to flow back and forth across the border between the two categories.

Lastly, it is important to consider the tension between critical approaches to broadly meta-stable assemblages of surveillance and government, with the potential desire to generate lines of flight. In effect, how can structures that theory holds to be relatively stable, and so understandable, be effectively challenged? If critical approaches generate lines of flight through these assemblages, practitioners can never be entirely certain that the assemblages will not coalesce into new, more troubling forms where the asymmetry of power is even more pronounced.

This is central to the issue of recursivity, and of the need to ensure a thorough approach to the assignation of causality. It is insufficient, as indicated above, to simply consider the relationship between the assemblage *as it exists* and the subjectivity of those surveilled. Instead, it is necessary to recurse to previous crystallisations of the assemblage to understand how the current assemblage emerged in its current form, and to examine the assemblants and their relationships that are consistent across those iterations. As outlined above, this does not mean a process of infinite regression, but rather a consistent challenge as to what the metastability of governments and policy means for the particular analytical case at hand.

Concluding Comments

The paper diverges from much of the established literature on methodology in Surveillance Studies in two main ways. Firstly, the paper considers that the subjectivist-objectivist debate that has characterised a significant portion of

the previous work (in one way or another) should be set aside, in favour of a critical realist (or critical *pragmatist*) view that is more congruent with the characteristics of both policy and surveillance systems (in their broadest sense).

Although public policy methods see widespread and daily use in the policy and surrounding communities, it is clear that they still lack, in most instances, the complexity to deal *adequately* with the rhizomatic and territorialising nature of surveillance. Similarly, governmentality offers significant critical potential, and has been used to good effect within Surveillance Studies, but still requires additional practical tools for the policy-maker to use it effectively. Adopting a 'realist governmentality', as a synthesis of the two approaches above means that attention can be paid to the difficult and contested realities; the multiple voices that are present within and around government.

Adopting an approach centred on assemblage theory allows some potentially new methods. The challenge faced at this stage is twofold: finding methodologies by which relationships between assemblants in any given coalescent structure can be mapped, and then finding a way to analyse the properties of those relationships. Understanding what the vector space or map of a particular assemblage tells us is its own challenge; understanding what several assemblages, which may be arranged in a causal chain, tell us is an entirely different order of magnitude, but one that should be approached with alacrity. The potential for critical insight, and practical impact on policy is extensive, and can allow Surveillance Studies to keep pace with the rapid change that both state and corporate surveillance assemblages are undergoing.

Part Three:

Research Approach, Fieldwork and Analysis

Chapter 6:

Original Research Approach and Methodology

Introduction

While the previous section examined the theoretical and epistemological considerations of the research process, this section recounts the practical design, implementation and operation of the research process. It also examines the 'real-world' application of assemblage theory to the research process. During the building and implementation of the research methods, a number of ethical considerations were raised.

The chapter lays out the original research approach, highlights the difficulties that were encountered in realising this approach, and the ethical protections that were put in place. In light of the difficulties faced gathering data under the original approach, the subsequent chapter then lays out how these problems were mitigated, by shifting to an analysis of online data through social media. As the main methodology that was used in the final version of the research, the full background and rationale for the selection of the topic of fieldwork is also detailed in that chapter.

Research Objectives

It is useful at this point to examine the research questions initially identified for the study, and how they were modified by the work in the subsequent chapter:

1. What are the policy assemblages relating to surveillance around the protest movements during 2010-12?
2. What are the components [assemblants] of those assemblages and in what circumstances do they persist or replicate to different contexts?

3. What strategies do the ‘pair actors’ use to enhance the likelihood of ‘favourable’ crystallisation of assemblages?
4. What is the effect of public policy assemblages on policy-makers and on the subjects of surveillance?
5. How can we utilise existing social theory to adequately explain and give insight into the operation of assemblages?
6. What does operationalising the assemblage in this way have to offer in terms of critical understanding?

The original intent was to both understand what a policy assemblage might look like, given a specific instance. Given this supposed discovery of the detail of an assemblage, it was intended that it would be able to trace a path of dependencies through differing assemblages, from policy through to lived experience, and understand how higher level assemblages, of policy and governance, crystallised around particular decisions, structures and praxis. The validity of the original questions is examined here, with a detailed rationale for the shift to analysing online discourse laid out in Chapter Seven.

With hindsight, the concept of asking “What are the policy assemblages?” was a malformed question. As has been demonstrated in the preceding chapter, assemblages tend to defy a singular description, being defined instead by their relationships, both interior and exterior. This is reflected in the change to the first research question focusing instead on understanding how *discursive* assemblages, which can arguably be tracked and analysed, persist and replicate across differing contexts.

The second original question survives in only slightly modified form, as it focuses on a crucial aspect of assemblages, the characteristics of the assemblage and their assemblants, and how the composition of an assemblage changes over time.

The third was posited on an assumption that drawing out details of some assemblage would allow policy intent and discourse to become available to the researcher. In turn, this would facilitate analysis of the policy process

from the perspective of the assemblage - how actors within those structures manipulate and manage the environment around to reach outcomes that shape the new structures in their favour. This was strongly related to the fourth question, which would have attempted to trace the relationship between those outcomes for policymakers, and the assemblages of surveillance that surround surveillance subjects.

Question five was retained in only slightly modified format, and attempts to make analogies from the existing field of literature in policy studies, surveillance and social theory to an assemblage based theoretical framework, with the groundwork for this laid out in Chapter Five.

Similarly, the final question was also retained unchanged, and offers an opportunity to explore the potential for critical work using this framework, as well as potential practical approaches to fieldwork.

Literature Review

The first stage in developing the research was to undertake a literature review. The outputs of this literature review have been laid out in Chapters Two through Five. The literature review took the form of an inductive and recursive review of literature. As this research is aimed at generating new theories or approaches, and is also an inherently multi-disciplinary study, there was no central hypothesis that could be directly tested.

Theories of surveillance were heavily directed by a series of main papers and texts, notably Haggerty and Ericson's *The Surveillant Assemblage (2000).* Using the JSTOR and EBSCOHost, papers which cited this were then explored and a list of journal articles was assembled. These journal abstracts were then examined for theoretical explorations of surveillance, rather than more practically-focused papers examining a particular instantiation of surveillance. This process was similarly followed using Lyon's *Theorizing Surveillance: The Panopticon and Beyond* (2006), using the papers cited and keyword summaries from chapters as a jumping-off point for further

exploration. This iterative process of reading, annotating and exploring further papers lasted several months. Crucially, this included Bogard's '*Surveillance Assemblages and Lines of Flight'* (Bogard, 2006), which opened a new avenue of research during the literature review, leading to further work on assemblage theories and theories of complexity, particularly the work of DeLanda, including *A Thousand Years of Non-Linear History (1997)* and *A New Philosophy of Society: Assemblage Theory and Social Complexity* (2006).

In terms of understanding the theory of assemblages laid out in later works, research began with Deleuze and Guattari's A Thousand Plateaus: Capitalism and Schizophrenia (1987), accompanied by the Cambridge Companion to Deleuze. The latter provided useful context and frameworks for considering this complex and notoriously difficult text. A journal search for 'Deleuze AND Complexity' was undertaken, which lead to Protevi's (2006) work on emergence and indeed to Deleuze's own Postscript on the Societies of Control (1992), which cross-references and relates strongly to the DeLanda work on assemblages and complexity in society.

Study Bounds

The study is both temporally and geographically restricted as it relates to protest movements and activities that have taken place since the 2010 General Election, and in the UK, consistent with the details of protests and the rise of digital activism in Chapter Seven. Specifically, and in practical terms, those demonstrations and activist movements predominantly took place in Greater London. Part of the hypothesis of the study is that it is possible and feasible to trace the relationship between policy and policy-as-experienced, and restricting fieldwork within these bounds will help to test this theory, by reducing the number of variables that may constitute assemblant components.

Original Fieldwork Methodological Approaches

Initial empirical research design took the form of a tripartite case study, utilising both interview and discourse analysis techniques. The research attempts to provide critical insight into surveillance policy by identifying and mapping the public policy assemblages in place around protest groups in the UK in 2010-12: understanding the impact of surveillance policy also involves documenting and understanding the subjectivity of the surveillance subject.

However, as noted, in light of the methodological and access challenges posted while conducting this fieldwork, the research shifted to examine protests within the UK in 2015.

Methodology

It was originally envisaged that fieldwork would be structured as a case study, an investigative tool used to:

> *"thoroughly describe complex phenomena, such as recent events, important issues, or programs, in ways to unearth new and deeper understanding of these phenomena."* (Moore, Lapan and Quartaroli, 2005).

Thomas (2011) defines a case study as the:

> *"...analyses of persons, events, decisions, periods, projects, policies, institutions, or other systems that are studied holistically by one or more methods. The case that is the subject of the inquiry will be an instance of a class of phenomena that provides an analytical frame — an object — within which the study is conducted and which the case illuminates and explicates."*

As explored in the policy process literature review laid out in Chapter Two, it is suggested that the full complexity of the public policy environment, particularly as it relates to surveillance, is rarely caught, so a case study is the ideal

vehicle for attempting to capture such complexity. Lapan and Armfield note that many different purposes for case study research have been identified in the literature, including its ability to explain, explore, describe and compare educational or social programs (Yin, 2003), and "to discover and communicate innovative ideas and programs"(Lapan and Armfield, 2009). In particular, the study was intended to be an *instrumental case study* under Stake's typology (Stake, 1995). The study uses case results to support the theoretical framework that it advances by applying it in the field and checking the validity. It is expected that the results will modify the framework in some way, and will be part of a process of clarification and modification, potentially over several such studies.

The intended outcome was comparative to process tracing within International Relations study, which attempts to highlight the causal factor – or factors – that lead to a particular outcome. Collier (2011) sets it out as "*the systematic examination of diagnostic evidence selected and analysed in light of research questions posed by the investigator*", particularly around Causal Process Observations. However, with the focus on overlapping and competing assemblages of decision-making that this study takes, a *single* decision point for any decision is not achievable or desirable. What is of interest is the relationship between a *series* of decisions, and their relationship to the lived experience of the policy.

The object of the case study was the ongoing and increasing surveillance of protests and dissent within the UK, while the specific subject was the anti-austerity protests from 2010-12, particularly the student protests of November 2010, and the Occupy London movement in late 2011/12.

To attempt to fully answer the research questions, the original study would be split into three parts, using two predominant approaches. The first two sections were interview led, focusing on the participants in protest and surveillance. The last section would be a discourse analysis of media and policy documentation.

Interview Sampling

Although any study must necessarily have regard to who is knowledgeable about the topic and 'willing to chat' (Moore, Lapan and Quaratoli, 2011) this study may raise particular issues around the selection of 'informants'. In the first instance, there is an issue of confidentiality and trust, particularly for those in the 'activist' community, who may have partaken in activity that was, if not directly illegal, then clearly and by definition, the subject of interest from the police and potentially the security forces. There is also the issue that those who are 'willing to talk' may have an agenda or particular viewpoint that they wish to convey, for their own purposes. With these in mind, the researcher has two options: the study can either look to undertake 'mass' interviewing, so as to go towards creating a representative sample such as might be found in more quantitative methods, or, the study can undertake a purposefully sampled approach.

Random sampling, or sampling the entirety of the available population is subject to several constraints and drawbacks: it does not necessarily eliminate the issue of bias, although it may make triangulating bias in some participants easier; it is subject to capacity issues, and may be subject to a high level of redundancy in the information gathered; although comprehensive, it does not give much, if any, statistical validity to the work.

Ethical Considerations

The research examines the practices and experiences of surveillance around people involved in political protest in the UK, 2010-14. The intent was to seek out individuals who have been under some degree of oversight and surveillance by the police and security services. It would be reasonable to posit that individuals who have a heightened awareness of surveillance methodologies and practices may have more experience and/or involvement within the protest movements or groups. As a consequence, there may be both a self-selection and a survivor bias in the sample of interviewees

available. This bias may manifest itself in a greater likelihood of having been involved in, or have awareness of illegal activity.

Through the course of the research it was possible that the interview process will uncover awareness of such activity. This raises two main ethical issues: potential identification of the individuals involved in such activity, and the awareness of ongoing illegality.

In the first instance, the researcher laid out a code of conduct for participants as to how their information, data and other associated details are used, including the right of withdrawal, to ensure informed consent to the research process. In addition to the normal statements of rights under UK Copyright law, this will also include pre-emptive statements as to when and where (such as under court order, however unlikely) the researcher may be compelled to provide information to the police or intelligence services. These statements form part of a consent form, which would be signed by the participant, with a participant number, and is held separately to the interview notes. The researcher will also provide copies of such notes as are taken (there is no intention to record interviews on any media) to ensure their veracity, accuracy and to provide the opportunity for participants to redact any potentially sensitive information after the fact.

In the second instance, on being made aware of ongoing illegal, or potentially illegal activity, the researcher would pause the interview to reiterate the legal position, and to make the participant aware that they are divulging potentially actionable material. The researcher will state that any disclosure relating to ongoing illegality that may result in immediate or proximate harm may be disclosed to police, but that the primary responsibility is to prevent harm (whether by exposure or by failure to expose) to the research participant.

Information retained is held in a secure environment, on a password protected and encrypted laptop. Backups are held on a remote platform, which is password protected at both login and individual file levels. This separation of

individual identifiers, personal data and password protection provides substantial protection from exposure, to all research participants.

Participant Selection

Three ‘gateway’ participants for the ‘activist’ interview were identified who may be willing to both participate in the case study and introduce the researcher to people who fit the following criteria:

- Aware of surveillance (whether Forward Intelligence Teams, or presence on National Extremism Database)
- Activist, involved in 10/10 protests or Occupy London
- Potential participants (names have been removed, as only one of these interviews took place at the time, and the material has not been utilised in the final form of analysis)
- Protester arrested during 2010 protests and subject of ‘police interest’ since that time
- Noted documentary maker/ activist
- Organiser of Plane Stupid protest group
- Associate at Bindmans Criminal Law (specialists in defending protestors and activists)

In order to capture similar detail on the policy and implementation side, one individual and a number of roles were identified as of interest. SCS indicates Senior Civil Service role classification, with SCS 3 being Director General level, and SCS1 being the most junior, Deputy Director.

1. DCS Adrian Tudway - National Co-ordinator for Domestic Extremism
2. NPOIU head and team members
3. Forward Intelligence Team (FIT) member(s)
4. CO11 (Met Public Order Unit) member(s)
5. Director of Crime (Home Office) (SCS2)
6. Gold Command for Student Protests
7. Director of Policing (Home Office) (SCS2)

a. Head of Public Order Unit & UK Football Policing Unit (SCS1)
b. Head of Police Transparency Unit (SCS1)

Even a cursory examination of Home Office responses to FOI requests indicates a high level of sensitivity, and no small degree of reticence about putting intelligence-related information in the public domain. The likelihood of gaining access to these primary sources is remote, which may fundamentally change the character of the research.

This research argues (both originally and after revision) that discourse is a key component of the acceptance and dissemination of assemblages. To achieve greater critical insight, the disparate levels of operation of 'framing' should be brought together with a consistent terminology. This should regard discourse not as isolated utterances or instances, but instead itself as an expressive part of a wider assemblage, allowing research to more effectively consider the relationship between policy and discourse.

Research was initially to take the form of three studies. These would examine the subjectivity of public policy around surveillance (and the subsequent impacts of that surveillance), from three perspectives:

- Individual (Occupy London/Wall Street protestors);
- Organisational (Metropolitan Police Territorial Support Group); and
- Policy (Government policy/media discourse) perspectives

Fieldwork would take two forms: semi-structured interviews and a media framing analysis, a subset of critical discourse analysis. Interviews would focus on a heavily surveilled group (Occupy London members and student protesters), and the directors of that surveillance (the Metropolitan Police's Open Source Intelligence Unit, and the Territorial Support Group, who operate the Forward Intelligence Teams seen gathering video and photographic data at large public events and protests.

Initial contact with protestors was through the researcher's own network, with two direct contacts, and one 'gatekeeper' contact in an activist group. It was intended that these multiple methods would allow a degree of triangulation (Easterby-Smith, 1991, Gill and Johnson 1991), and compensate for the weaknesses of singular methods. In particular, it was intended that while discourse analysis would provide a meso-level analysis, it would be weak on providing understanding of motivations and feelings at an individual level. This weakness would be undesirable in a study aiming to bridge the gap between assemblages of policy and those of the individual, and so a mixed-methods approach was taken.

Police access, particularly for early stage researchers, can be difficult. Without an established publication record, it is difficult for institutions to gauge a researcher's motivations, and in turn, how they are likely to treat any information uncovered. Lack of access means that the comparative analysis may need to move from individual/institutional to an international focus. Alternatively, the balance with other aspects of fieldwork may be restruck to create a suitable comparator.

The second piece of fieldwork was intended to be a content analysis of media and policy relating to surveillance, with the same focus on groups as above. Discourse constitutes an important element of the construction of both public policy and identities, but is often overlooked, or the existence of discourses are taken as ends in themselves (Barnard-Wills, 2012). Undertaking a discourse analysis would allow critical insight into the constructions of policy and identity that the interviewees relay, and is helpful in understanding the context and political economy of the protest and anti-protest forces. In particular, it is useful to approach the subject in this manner to attempt to develop a critical theory of surveillance public policy. This would provide a balance of academic critique and recommendations for policy-makers – or those resisting policy – about how the process is conceptualized and organized.

Initial interviews were scheduled with two activists who had been the subject of police scrutiny and interest after a protest in early 2011. These were 'scene-setting' interviews, to provide context and to ensure that subsequent interviews with activists could be handled sensitively and ethically. Taking a semi-structured approach allowed initial areas to be laid out as a grounding for the discussion, but simultaneously allowing the coverage of the interview to expand in a naturalistic manner (Douglas, 1986, Easterby-Smith Thorpe and Lowe, 1991). An initial phone interview was also held with the organizer of an activist network (Netpol, the Network for Police Monitoring) (not recorded or formally noted, and solely for scene-setting), and a request relayed to the activist network through that same organizer.

After initial interviews, feedback was provided to the researcher that the research was making those around the initial two interviewees deeply uncomfortable; the gatekeeper contact also advised that there was both disinterest in the research, and some degree of hostility towards an outsider, even though researcher access was being mediated through one of the group's leaders. In addition to the obvious ethical considerations of such feedback, it also had the direct effect of peer-pressuring the first interviewees into withdrawing their consent to participate in the study. Notes from the original meetings were destroyed at this point, in line with the original ethics submission.

Discourse Analysis

Discourse analysis is a crucial element of the research. This study argues that assemblages are made up of both material and expressive assemblants; some form of textual or linguistic analysis is an excellent method to understanding their composition, direction and sentiment, providing what Deleuze and Guattari (2004) termed the "*collective assemblage of enunciation*".

Barnard-Wills (2011:550) notes that this:

> *"also include[s] a capacity for mapping shifting links in the linguistic assemblage… and the insight that media surveillance discourses are neither monolithic but rather multiple and fragmented."*

Although media representations are, and will remain, important in this study, they will presumably not represent the entirety of the expressive assemblants, and so the research must also look to policy and related documentation, to help trace the 'ideational' elements that enter decision-making process, just as they would be tracked through in a process-racing approach.

It is the relation between each individual assemblant and the overall assemblage at any given point in time that constitutes the area of interest: what is the overall 'feeling' on the subject at any given time (sentiment analysis), and how does this particular assemblant relate to that whole? This piece of work has the potential to be a published piece of empirical work in its own right, and would be an original contribution to the field.

Sampling Method

The original approach was to make use of a pseudo quantitative approach, which undertakes a deep textual analysis of several key articles, and uses these to create coding frames for further data gathering around the topic. Selected publications (typically, but not limited to, mainstream newspapers and news magazines) are then searched for a series of keywords derived from the deep textual analysis. Presence of the keywords indicates initial inclusion, and the articles are then screened to ensure relevance.

To provide this study with sufficient depth, the time period selected for analysis was from the May 2010 election, when the possibility of protest was first raised, through to the end of 2013. This allows the inclusion of build-up, event occurrence, and then any fallout, 'post-mortem', and where appropriate, legal outcomes relating to the protests in 2010.

Initial media, from which the ‘key articles’ are drawn from five national newspapers: The Times, The Guardian, The Independent, the Daily Telegraph and the Daily Mail, along with their corresponding Sunday editions. These were selected on the broad assumptions that:

- they represent the centre, left and right of mainstream political discourse
- they represent a significant majority of the overall print media circulation in the UK, and almost the entirety of the non tabloid press
- their audiences are sufficiently diverse as to represent a high-level discourse between different sections of society

Table 1 - Circulation and Social Class Breakdowns for UK National Newspapers

	Total Circulation (000)	Social Class Breakdown	
		ABC1	C2DE
Daily Mail	4839	3198	1641
The Telegraph	1788	1587	201
The Times	1791	1587	204
The Independent	686	565	121
The Guardian	1264	1128	136

The textual analysis would be undertaken by examining the texts for frame elements, comprising three distinct areas. The first area is what could be called the 'mechanical' elements or attributes, the elements of the article that make it fall within the subject area at hand, and allow the other frame elements to operate. For this study, it would be reasonable to assume that 'mechanical elements' would comprise mention or discussion of:

- Occupy London
- UK Uncut; or
- Student protests or protesters; or
- 'Intelligence-led' policing relating to any of the above

The second set of frame elements could be termed the 'attitudinal' elements, which dictate the orientation of the article towards the mechanical elements – positive or negative, thematic or episodic coverage, individual or systemic attribution of responsibility (Semetko and Valkenburg, 2000).

The third set of elements comprise the relationship between the article and its predecessors. This study purports that discourse is most usefully conceptualized as a vector, that is to say that it has both magnitude and direction. Discourse and dialogue (as argued in Chapter 2) only reach signification *in relation to other fragments of discourse*. Without intertextuality (explicit or not), they are effectively meaningless utterances. The practicality of that for this study is that the relationship (if any) between two fragments can and should be mapped to provide a relative direction for each article or discourse fragment. Although the deep discourse analysis was abandoned in light of subsequent analysis, this concept of *vectors* of discourse is an important one that is resurfaced in light of the empirical work carried out.

Availability of Access and Reconsidering Approach

The Metropolitan Police were approached with a request to speak to officers in CO11, Public Order Operational Command and/or the Open Source Intelligence Unit, headed by Umut Ertogral. These requests were initially

made through the press office, and then through follow-ups with officers and named divisions from previous exchanges. Although initially promising, the request was ultimately met with obfuscation and confusion, with referrals across the service, and to various officers who either had no knowledge of the subject, were unaware of the request, or declined to return messages. Given the consistent stance of the Metropolitan Police Service towards access to this group[11], this is perhaps unsurprising.

Taken together, this removed two of the key aspects of the original research design; it also served as a warning that the research was over-broad and over-ambitious for a single researcher project at doctoral level.

Reflections on Failures of Access

In both cases, with activist groups and the Metropolitan Police Service, the failures were largely down to issues of credibility and legitimacy. As a relative unknown to the protest groups, and as a complete outsider to the police service, the researcher lacked any significant credibility with either group.

Issues with access to the protest and activist networks could potentially have been mitigated by a process of co-work and co-construction with the groups, earlier in the process. By offering transparency and a degree of direction in the subject matter, and working with them over a period of time, it may have been possible to build relationships inside the groups over a period of time. However, this was not accounted for in the early research process, which was largely archive and theoretically based. IN practice, this meant that the researcher had both limited time to conduct fieldwork in, as well as a very specific vision of the type of research, subject matter and shape that the interviews would take, that may not have been conducive to fomenting constructive relationships with a group who have significant reservations about ‘outsiders’.

11

From a police or policy official perspective, a similar issue arose. Without a record of publication that could be referred to, it would be difficult for officials and gatekeepers within the police service to judge the approach that the researcher would take in terms of the research's treatment of police actions. This came at a time when the police had been recently and heavily criticised for their actions around the G20 protests, Student Protests and London riots in 2011 by both the media and police standards bodies.

The lack of access could have potentially been addressed through the networks available to the researcher, through the Surveillance Studies Network, the Centre for Research in Information, Surveillance and Privacy. However, the time pressure that was identified in relation to building relationships with the protest groups was also at play in facilitating this access. The lack of engagement with potential sources earlier in the process must be regarded as a decision that had a fundamental effect on the availability of sources available to the researcher, and which drove the need to shift the focus of data gathering entirely.

Mitigation

The course of action to mitigate a lack of access will depend on the extent of that lack, but can be summarised in three general categories:

- Refocus the participant list – for example, to former or retired police officers, or to other policy officials within connected spheres
- Refocus the research to create a cross-country comparison (for example, between the Occupy London and Occupy Wall Street groups), or
- Attempt to expand the scope and/or depth of the discourse analysis to compensate for the lost policy and tactical insight that would come from these interviewees

Reconsidering Empirical Research

In light of the issues with operationalising this research, the original research questions were reconsidered, as discussed in the introduction to this chapter. After substantial discussion with the supervisory team as to the nature and purpose of the research, the research was reformulated to undertake a large-scale analysis and mapping of discourse around protests and surveillance. The value of such large-scale approaches has been subject to some debate (Mahrt and Scharkow, 2013), over whether to capture all data available, which is programmatically easy to do, or a more robust and typical sampling approach is more appropriate; these issues are explored in detail in the next chapter.

Chapter 7: Revised Approach and Methodology - Refocusing to Social Media Analysis

Introduction

As indicated in the introductory Chapter, the final research questions that the study attempts to answer are:

- What discursive assemblages are in place, and how do they persist or replicate across different contexts?
- What are the characteristics of those discursive assemblages and how do they change over time in relation to 'real-world' events?
- How can we utilise existing social theory to adequately explain and give insight into the operation of assemblages?
- What does operationalising the assemblage in this way have to offer in terms of critical understanding?

Refocusing – to Twitter data

Twitter is a 'micro-blogging' service, where users post tweets: messages of no more than 140 characters[12]. Twitter has undergone several amendments to its core functionality, which enhance the 140 character limit, for example allowing links to be shortened and embedded automatically, and capturing message 'threads' by showing replies to tweets in sequence. Crucially, it also has an Application Programming Interface (API), which allows researchers to develop their own systems to capture and analyse data. Such APIs also form part of the basis of Social Media monitoring systems (SOCMINT) as used by the police and intelligence agencies (Omand, Bartlett and Miller, 2012).

What was the decision to refocus to public twitter data based on? Further risk mitigation – known level of data and that it would be available (barring a disastrous platform loss) – lost time and issue of general uncertainty.

What would Twitter data provide? Information about the discourse at protests – in itself a node and topic of a case study – there were at least two

12

suitable protests known about which would put us back to towards the right level of case study/volumes.

UK Protest and the Worldwide Shift to Networked Activism

In the wake of the 2007-08 worldwide financial crisis, the UK has experienced an ongoing period of wage stagnation and economic shifts, notably an ongoing and significant recession between early 2008 and the beginning of 2009, reflecting similar worldwide conditions. Against this background, a Conservative-Liberal Democrat coalition government was elected in the UK in 2010, which implemented a policy of austerity, cutting public sector expenditure significantly, ostensibly as a response to the financial crisis and subsequent gaps in public sector finances.

In 2009, partly as a response to the nascent crisis, there were widespread protests in London against the G20[13] meeting being held there in April, including protests at the Excel Centre in East London, a Stop the War march in central London, and extensive protests around the City of London and the Bank of England[14]. The 2009 protests were notable for their critical coverage of the police response to disorder and public order incidents, as well as noting the rise in public awareness of the police practice of kettling protestors[15]. A protestor, Iain Ttomlinson, died as a result of a police baton strike at the protests whilst walking away from police lines, and was later ruled to have been unlawfully killed. The HM Inspectorate of Constabulary, the police review body at the time, reported on the G20 protests and noted the sophistication and speed with which protest groups were able to organise and regroup by using online technology, such as social media and mobile phones. (HMIC, 2009, p25).

The G20 demonstrations were followed in December 2010 by a number of student protests against the then-proposed changes to higher education

13

14

15

funding by the Conservative-Liberal Democrat coalition, elected in May of that year. The protests again saw large scale containment and kettling of protestors, and violent clashes between police and protestors, with accusations of widespread and disproportionate violence on the part of the press. The protest was particularly notable for the violent occupation of Conservative Central HQ on Millbank, and the hospitalisation of Alfie Meadows, struck in the head by a police baton (Guardian, 2010)

Beyond the UK, there were large scale protests, spilling into civil resistance, beginning in Tunisia in late 2010 with a month long protest that became a resistance and opposition to the Ben Ali government, which ultimately collapsed. Triggered by the self immolation of Mohamed Bouaziz in the town of Sidi Bouzid, media analysis of the protests focused on the role of the internet and social media in facilitating activist communication. At one stage, access to Facebook was restricted to such a degree that only around 20% of the population had access, and there was a state-level man-in-the-middle[16] attack on Facebook passwords during the uprising[17]. Although the level of facilitation through the internet has come into question, the initial trigger for the protests was the recording and uploading to Youtube and Dailymotion of protests immediately following Bouaziz's death (Castells, 2012).

In 2011, there were widespread protests across Oman, Yemen, Egypt, Syria and Morocco, with Egyptian President Hosni Mubarak resigning in early February and installing a military council to rule the country in the interim. Ultimately, the protests in Syria and Libya spiralled into civil war after the state responses violently suppressed the civil uprising and activism. The Arab Spring, as it became known, is a large topic worthy of its own extensive literature, but is presented here as background to a rising wave of activism enabled, facilitated and organised by online media and tools. A timeline of notable protests, including those known to have been influenced by social media, is included at Appendix 5.

16

17

Returning to the UK, on March 27 2011 the TUC held a large protest march to demonstrate against spending cuts that that coalition government, elected in 2010, were implementing. Attended by around 250,000 people, the protest was the largest single demonstration since the anti-war rally before the second Iraq war in 2003. [18] The police viewed that there were "*...approximately 500 criminals committing some disorder including throwing paint at Topshop in Oxford Street, and at police…*" (BBC News, 2011).

In late summer 2011 there were a series of riots across England, primarily focused in London, which were triggered by a continued escalation of responses following the police shooting of Mark Duggan on August 4th. Duggan's shooting by Metropolitan Police was eventually ruled lawful after a public inquest, but clashes between locals in the Tottenham area of London and the police escalated into widespread public disorder in, at first, Tottenham, then other areas of London, before spreading to other cities in England.

The causes, triggers and proceedings of the riots are themselves subject to significant academic attention, but the 2011 riots are again notable for the widespread attribution and blame placed towards social media, in this case, BlackBerry Messaging (BBM) in the spread and organisation of participants in the disorder. RIM, the manufacturer of Blackberry smartphones and provider of the BBM service were known to assist police under the Regulation of Investigatory Powers Act.[19]

In the United States, the Occupy Wall Street movement began in Autumn 2011 with a series of posts on the Adbusters website, an anti-consumerist not-for-profit and pro-environment publication based in Canada. In June 2011:

> *Then, in early June, the art department designed a poster showing a ballerina poised on the "Charging Bull" sculpture, near Wall Street.*

18

19

> ...*"the juxtaposition of the capitalist dynamism of the bull," he remembers, "with the Zen stillness of the ballerina." In the background, protesters were emerging from a cloud of tear gas..."What is our one demand?" the poster asked. "Occupy Wall Street. Bring tent."*
> (Schwartz, 2011)

Adbusters sent an email to their subscribers, suggesting that "*America needs its own Tahrir*", a reference to the epicentre of the Egyptian protests during the Arab Spring. (Schwartz, 2011). In spite of the suggestion's early traction on Reddit and other Twitter, one activist ('P') noted that:

> "If we'd used a mass text message, or Twitter, it would have been easy for the police to track down who was doing this." (*ibid.)*

After agreeing a manifesto through a general assembly, the Occupy Wall Street campaign held Zuccotti Park in New York for 59 days, until being forcibly cleared by the New York Police Department on the 15th of November, with around 200 arrests taking place, including members of the press and legal observers.

Occupy London was an ongoing protest movement and series of events between October 2011 and June 2012, in response to the main Occupy Wall Street movement, which had already spawned a number of similar movements in the US. As with other protest movements, the trigger for the commencement and organisation of Occupy London came via social media. In early October, a series of Facebook posts were made, calling for protests at the London Stock Exchange and the Bank of England, in solidarity with the Occupy Wall Street protests in the USA.

In the UK, ongoing protests against austerity and budget cuts saw 150,000 attend a Trades Union Congress march in London in October 2012, with matching protests in Glasgow and Belfast (BBC, 2012).

During June 2013, there was increasing disquiet at the actions of the Morsi government in Egypt. This culminated in widespread protests nationwide, and

centred again on Tahrir square in Cairo in June, marking the one year anniversary of Morsi's inauguration. Military estimates from aerial reconnaissance of the crowds estimated that 14 million people across Egypt took part in the protests. Ultimately, the protests led to a *coup d'etat* by the Egyptian military in July of the same year.

In 2014, further anti-austerity protests were seen in the UK around the People's Assembly gatherings and meetings in London in June, while in Hong Kong, large scale occupations, protests and civil disorder were seen in September, as the 'Umbrella Movement', named for the umbrellas used by activists for defence against police pepper spray.

Background on Surveillance in the UK

The United Kingdom has a long history of surveillance, particularly in terms of surveillance cameras, and monitoring of protests and activists. A full history of UK surveillance could (and does) fill numerous volumes. However, some specific examples are useful to contextualise the choice of topic for this research, in conjunction with the sections on the rise of social media and the upturn in protest and digital activism in the first part of the 2010s. Additionally, the UK's participation in large scale internet dragnets has come to light, with systems paralleling the US National Security Agency's monitoring strategies, as well as continuing revelations about the extent to which undercover police officers have infiltrated the lives of activists.

The UK is something of a world leader in the adoption of the closed circuit television cameras (CCTV), used to monitor a variety of locations, from city streets and train stations, to private use to monitor building interiors and prevent shoplifting. The proliferation of CCTV systems in the UK has been particularly notable since the 1990s (Webster, 2009), with some arguing that that the UK is amongst the most surveilled nations in the world (Norris, 2004). Webster argues (Webster, 2009:21) that increased awareness of CCTV and their associated effects (or lack thereof) could lead to a public pushback

against the extension and promulgation of these systems. However, this has yet to manifest itself in any significant form, with demonstrably low public levels of awareness of the surveillance that surrounds the average UK. (Renaud, et. al., 2016).

Coleman and Sim (1998) argue that existing literature on CCTV has largely underrepresented the role of an authoritarian state in the roll out of camera systems. This research argues that while there is a case to be made that emanations of the state have played a significant part in the proliferation of surveillance technologies, positioning this issue in terms of a centralised and singular authoritarian state is unhelpful. Instead, the relationships between the assemblants that make up that state, and why they find it useful in terms of overlapping (and sometimes competing) goals to 'militarise' public space (Davis, 1990, 1992) is a more nuanced approach.

Similarly, there is a historical precedent and practice of placing undercover police and intelligence officers within activist groups. While this practice dates back to the earliest days of Special Branch in the 19th Century (Porter, 1991), there is an increasing tendency for this to be focused on non-violent domestic groups protesting corporate issues. Such undercover activities, with state agencies including police forces and the intelligence services investigating UK citizens protesting against corporate issues highlights the increasingly blurred delineation between public and private issues, and between the state and corporate concerns. (Lubbers, 2015). This was brought into the public consciousness with the details of the Mark Kennedy 'spycops' case, where an undercover officer engaged in lengthy and intimate relationships with the persons under investigation (Walker and Kingsley, 2014).

The UK has placed an increasing focus on social media monitoring in recent years, ostensibly for the policing of 'domestic extremism', (Dencik, et. al., 2015) much as previous practices of embedding undercover operatives in activist and potential terrorist groups did. This utilises Open Source Intelligence (OSINT, also referred to as Social Media Intelligence, SOCMINT) to collect publicly available data without bypassing privacy settings, or having

to overcome encryption of communications. These strategies are used to inform approaches including pre-emptive arrest and police tactics during and around protest events. (ibid).

The 2011 riots in London and across England saw the police and government raise the issue of the difficulties in monitoring encrypted messaging services. This desire to include private messaging systems in the toolkit of law enforcement and intelligence services has persisted throughout the decade. The UK Government making repeated requests to RIM (Guardian, 2011), Facebook (Guardian, 2014) and Twitter (2019) for data, as well as repeatedly arguing that encrypted messaging services should provide the 'keys' to their services to allow intercepted messages to be decrypted (The Verge, 2017).

In addition, the UK's Tempora program, an analogue of the United States National Security Agency's PRISM program, reportedly placed over two hundred interception devices on telecommunications cables connecting the UK to Europe and the United States. These devices gather telecommunications data on an 'industrial' scale (Bauman et. al., 2014). Such mass surveillance contains a strong emphasis on categorical suspicion; rather than being closely targeted, it instead opts to collect data on almost all communications through these channels, and are then organised through programmatic platforms that allow visualisation of the networks of communication.

The Emergence of Twitter as Locus for Public discussion

Neumayer and Rossi (2016) noted that media technologies are not 'clearly identifiable machinery', but instead are semantically rich *concepts* (as well as rich data sources) that are closely interwoven with societal expectations and behaviours. In many respects, the 'platform' of social media is not clearly identifier as medium or message, instead the two are closely interwoven, with the methods that technology companies use to curate and manage the content on their platforms shaping debate in ways that are still to be fully

understood. However, in terms of shaping fieldwork for research, it is important to understand Twitter's prominence in online discourse over the period of research.

Networked activism emerges as a tool to be studied in late 2000s (*ibid*). The uptake and penetration of social media as a whole moves the concept of studying networked activism and discussions on social media from a niche constituency to something more reflective of overall population.

Protest and Surveillance in the Age of Social Media

The preceding sections indicate not just an upturn in the number and level of protests and civil disturbances around the world, but also a confluence between a number of factors that are of significant academic interest. The first is the rise in protests coinciding with the increased availability of social media. The second is the nature of the data that can be gathered from social media from an academic perspective: inherently discourse based, semantically rich, real-time, publicly available and highly structured, all of which makes social media attractive as a basis for study. The final factor is the use of those social media technologies by groups to resist structures of power and undertake their own organisation and mobilisation, using the very processes and technology that surveil them.

Social Media Growth

Social media grew exponentially in between the mid 2000s and the time of study. From a niche interest in the mid-to-late part of that decade, social media sites and platforms now count a significant portion of the world's population, across all countries, as their user base. The most useful metrics to understand the relative reach of these platforms is monthly active users (MAU), which counts the number of users who have logged in within the last 30 days for any given period.

Facebook is simply the largest social media and content platform in the world, with an unparalleled reach. Steady linear growth across the decade has taken Facebook from 431m monthly average users in Q1 2010 to 2.32bn in Q4 of 2018. Although an order of magnitude lower, Twitter has seen similar rises in usage, from 30m in Q1 2010 to 321m in Q4 of 2018. While Facebook's user base expanded steadily throughout the decade, Twitter's user growth largely stalled in 2015, with the platform reaching 302m monthly users in the 1st quarter of that year[20].

The UK represents a significant portion of Twitter's approximately 320m user accounts, with the UK the fourth largest geographic user in the world after the USA, Japan and Russia. Further, the 17.1m (in 2018) users[21] in the UK represents a significant proportion (25.9%) of the UK's entire population) . At the time of fieldwork, the UK user base was 14.8m against a population of 65.1m (22.7%). Although the age requirement for a twitter account is 13, there is a reasonable degree of overlap between the Twitter population and the working age population (defined by the ONS as 16-65). In 2015, the working age population was 63.3m[22], giving a market penetration for twitter of 23.3% with an unfortunately unknown margin for error around the overlap of people aged 13-15 who can legitimately sign up for an account, and around those who may have signed up before they were allowed to do so under the platform's terms of service.

Programmatic Research

Social networking sites such as Twitter can provide significant levels of data for researchers. However, they can lack contextualized information gathered through more ethnographic methods such as interviews, or embedded digital ethnography (Hand, 2014).

20

21

22

In an attempt to capture the overall assemblage (or as more typically referred to in social network studies, the *graph*), the research design was altered to capture ongoing thematically relevant content from Twitter. As graph terminology is used in Chapter 7, it is appropriate to offer a definition that a graph is*:*

> *"[A] mathematical structure consisting of two finite sets V and E. The elements of V are called vertices (or nodes), and the elements of E are called edges. Each edge has a set of one or two vertices associated to it, which are called its endpoints"* (Gross and Yellen, 2005:2).

Graph theory has the potential to extend and complement the more general and ambiguous assemblage theory, if values are assigned (textual, numeric or otherwise) into those assemblages, although there are certain limitations on the knowability of, for example, personal assemblages of thought and experience.

The refocused research was intended to remain retrospective, examining the student tuition fee protests in 2010, which are already well-documented. However, archived Twitter data is not publicly available, as it is one of the income streams for the company, and is commercialized through partners such as GNip. A quote was requested for a relatively small data sample (around 500,000 tweets, targeted by keyword and date-range), and was in the region of £7,000; further refinements of this sample would be at additional cost. For this reason, it was decided to use a future event, and prepare software that could capture and analyse contemporary data as it was generated.

During background research, the manual search functionality was used from the Twitter 'developer panel', which is designed to allow early exploration of the data available to developers and researchers.[23] This retrieves a sampled set of recent tweets containing the search term. However, this was not

23

suitable for development into a programmatic solution for a number of reasons:

- Twitter's sampling methodology (and therefore potential biases) are proprietary, unclear and therefore an unknown in terms of data structures;
- Samples are only available for the last 7 days;
- Samples are limited to 100 tweets per request;
- Twitter implements 'rate-limiting' on searches, meaning that individual applications can only make a certain number of requests per day (around 17,000), limiting total tweets per day to around 170,000; (Twitter receives on average about 347,000 tweets per minute globally[24], and in the UK has seen tweet rates above 22/second on specific topics, such as the 2010 and 2015 general elections. (Semiocast, 2010))
- Overlapping searches (e.g. separate searches for "MI5" and "Surveillance" would both match "Snowden confirms MI5 surveillance program", but would still count towards the rate and tweet limits, reducing the effective number of tweets that could be gathered for analysis.

Cesare, Grant and Nsoesie (2016) note incorrectly the issue of rate limits for Twitter:

> *"...Twitter's public API… includes measures designed to slow the collection of high volumes of data… most calls... are limited to 180 calls per 15 minutes. This means that only 180 users' metadata can be gathered within 15 minutes… to access data on users' ties, one must first pull the Ids of friends and followers and then link metadata to these IDs – a time consuming, two step process."*

24

Although this is true in terms of the rate limits themselves, each rate limit call will allow the retrieval of 100 users (or tweets) per request. In practice, it is possible to extract user metadata from Twitter at the rate of 72,000 users per hour, (over 1.7m per day) by using the free API. Rate limits for research can be increased or obviated entirely by purchasing a premium tier package from the company. As an example, presently, these start at $149/month for a 500 tweets-per-request / 60 requests per minute package, increasing the potential volume of data available by a factor of 25.

Working within the confines of the free tier, the program was rebuilt to consume the Twitter 'stream'. Rather than 'polling' the Twitter servers a number of times per second, this approach opens a permanent connection to the servers, with a pre-specified list of parameters. Twitter then serves a sample of any matching tweets, for the duration of the connection. However, the number of tweets returned is entirely dependent on the volume of matching tweets generated; as this process is effectively real-time, it is also effectively uncontrollable.

There are several tiers of access to Tweet data, with the Twitter 'firehose' at the highest level, and individual keyword searches at the bottom. The Twitter firehose represents a syndicated data feed of the entirety of content being posted on Twitter, in real-time. It represents an enormous volume and velocity of data, and few organisations use it in its entirety. Some may opt to use the enterprise 'Decahose stream', which is a 10% random sample of the Firehose (Twitter, 2017). A further moderated level of access is to the Twitter 'stream', which is a further sampled version of the firehose, and which provides a maximum of 1% of all 'current' tweets as they happen. If the number of tweets for a particular keyword or keyword set exceeds 1% of all current tweets, the dataset is 'down-sampled' to fit that limit (Morstatter et al, 2013). Morstatter *et al.* also note that the down-sampling effect of moving from the firehose to a sampled stream means that for a term n, the sample of keywords and hashtags is broadly reflective of the larger population where the population size is large, but can significantly under-represent keywords where

the population is smaller. They also note that , in particular, "*topical analysis is most accurate when we get more data from the Streaming API*" (Morestatter et al, 2013:406). However, they do go on to note that:

> *'...surprisingly… the Streaming API almost returns the complete set of the geotagged tweets despite sampling… researchers using this information can be confident that they work with an almost complete sample of Twitter data when geographic boundary boxes are used for data collection." (Ibid:407)*

However, as noted in the analysis section in Chapter Eight, not all tweets contain geographic metadata; some users will have turned off, or never enabled this functionality, while those who are using the service from a desktop computer have to undertake a number of additional steps to add this data to their tweets. The default for users on mobile devices such as smartphones and tablets is for this to be enabled, but the availability of geographic location is dependent on the devices location services (GPS) having been turned on, and permission given to the application to access it, which is far from certain in all cases.

Programmatic Research Ethics

As the research design was altered in response to the demands and challenges of the original design, the focus of ethical concerns shifted. The primary concern of any research must be to do no harm to the participants or observed groups in the research. Central to ensuring that requirement is obtaining informed consent. How this operates at scale, in online communities, can be significantly different from considering the proximate potential for harm to an interview subject, that the researcher builds up a rapport and trust with, over several encounters.

However, the utility and richness of data from online communities has been recognised since early in the web's lifespan (BMJ 2001;323:1103). The rate

of online data created and interactions between people grew exponentially with the beginning of the 'Web 2.0' era, which focused on 'user-generated content', rather than content solely managed and provided by website owners. (O'Reilly, 2007) Facebook, and later Twitter and instagram gave most people (in the West) an obvious online 'home' by 2010. An additional factor to consider in whether or not individuals have given effective informed consent to their data being used is that Twitter offers 'protected' accounts, where other users must request access to the protected account before tweets will appear in their timeline. Protected accounts and their tweets do not show in the timeline, unless the accounts are already followed by the owned account.

In 2012, a research team working in Facebook, and supported by academics from Cornell University in the United States, actively manipulated the News Feed content of 700,000 users' Facebook pages to induce changes in their emotional state, and which finally came to light in 2014 when the findings were summarised in the Proceedings of the National Academy of Sciences (Kramer, Guillory and Hancock, 2014). This paper was held widely as an example of bad practice (Panger, 2015), both in terms of ethical practice and regard for the research subjects' well-being. Users had no information that they were, or even may have been, involved in an experiment, and were given no opportunity to consent, on an informed basis or otherwise. Similarly, although the study was ostensibly given to investigate whether users' emotions were *negatively* impacted by seeing positive content on others' profiles, rather than to directly manipulate people into a negative emotional state, it had very low internal validity (Panger, 2015:1111), and may still have a negative effect on users' emotional well-being, an outcome that is not justified by any means by the aims of the study.

Informed consent is underpinned by three core principles: Respect for autonomy, beneficence, and justice (Faden and Beauchamp, 1986). Although the US Federal guidance Common Rule (DHHS, revised 2009) does not apply in the United Kingdom, it provides a useful baseline for research involving behavioural research, §46.111 providing that:

> *"Informed consent will be sought from each prospective subject or the subject's legally authorized representative, in accordance with, and to the extent required by §46.116."*

§46.116 of the same guidance provides that the subject must be informed of the purposes, risks and benefits, alternatives and how many other people are involved in the study. However, it does also provide for a waiver to informed consent (§116.d):

> "Where *the research involves no more than minimal risk, the waiver does not affect the rights of the subject, the research could not be reasonably carried out without the waiver, and the subjects will be informed, wherever pertinent*."

Eysenbach and Till (2001) note that although there is an analogy between placing a post on the internet and a 'letter to the editor' publication, there may be an underlying psychological difference, depending on whether the group or community is relatively closed (as the Usenet groups considered by Eysenbach and Till were), or more open. This research strongly argues that although there may have been an expectation of 'not being a research subject' online at the time of writing (2001), the widespread apathy in the face of repeated revelations about *secret* surveillance means that expectations may have changed, in line with Nissenbaum's principle of contextual privacy (Nissenbaum, 2011), particularly given the increasingly commercial scale and focus of web platforms and technologies. Assuming that data that is public or semi-public and may be re-used does not, of course, constitute informed consent, but it does arguably change the dynamic and balance on what consent needs to be sought from people who are caught up in a mass observation, particularly in a society where public posts on Twitter are frequently reused (often without consent) in articles in newspapers, blogs and online magazines.

The Association of Internet Researchers (AoIR) offers some guiding principles that should be considered (AoIR, 2012) while pivoting the research methodology to an internet-based piece of programmatic research:

- The more vulnerable the community or participant, the greater the obligation on the researcher to protect them;
- Harm is contextual and ethical principles should be understood inductively rather than applying universally;
- All digital information involves persons at some stage and so guidelines on protecting human subjects should be consulted, even if not immediately obvious how this applies in context;
- Researchers must balance the rights of subjects against the social benefits and costs of research - the rights of the subject may outweigh the rights of the researcher, or the benefit of the research;
- Ethical issues should be considered during every step of the research process; and
- Ethical decision-making is a deliberative process, which should involve as many people and resources as possible.

Following the question structure laid out in the AoIR guidelines, the following can be stated about the structure and intent of the research:

- *How are we recognizing the autonomy of others and acknowledging that they are of equal worth to ourselves and should be treated so?*
    - The fieldwork is purely observational, with no interaction between researcher and subjects, so no impingement on autonomy. Subjects' rights to privacy will not be harmed, as no specific data will be used in any identifiable way – the study focuses on metadata and meso level analysis – no individual tweets or individual data will be surfaced in the analysis and presentation.
- *What are the potential harms or risks associated with this study?*

    - Age is an issue – as discussed in Chapter Eight, this information is not explicitly provided on Twitter, and there is no obvious or practical way to scan for minors. However, as indicated, the study will not surface specific text or content.
- *How is the context (venue/participants/data) being accessed?*
    - Programmatically, through an Application Programming interface provided by Twitter, which provides access to their data in near real time.
- *How are participants / authors situated in the context?*
    - Participants in a social media platform, sharing content and discussing issues.
- *How are participants/authors approached by the researcher?*
    - Not approached directly. Researcher is solely observational; users have agreed to reuse of their publicly available data when signing up for the platform by agreeing to the Terms and Conditions.
- *How is the researcher situated in the context?*
    - Solely observational. Removed from immediate context as data is gathered in real time and analysed significantly after the fact.
- *If access to an online context, is [the resource] publicly available?*
    - Yes
- *Who is involved? What is the primary focus of the study?*
    - General population, with a subset based on keywords, of people who are discussing the research subject.
- *What particular issues might arise around the issue of minors or vulnerable persons?*
    - Significant issues, if any specific or identifiable information was to be used in analysis or publication, but only meta/meso-level

aggregates are used. There is no way to re-identify individuals from these aggregates.

- *What are the reasonably "foreseeable risks or discomforts"?*
    - Identification of an actionable, objectionable or otherwise notable statement, in a manner that affects the maker; in this particular instance any identification of a social media user should be regarded as negative. Although the expectation of privacy is highly contextual and will vary across users, researchers should seek to avoid any impact on individual users.

The principle of Value Sensitive Design (VSD) is central to how this system and research project operate. Albrechtslund (2007) defines VSD as:

> *"...a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process."*

Albrechtslund (2007) examined the VSD principle and concluded that design intent and user practice often do not match, and the relation between designed intent and eventual use is inherently complex and unpredictable, and instead promotes a more phenomenological approach to understanding the interaction between users and their systems. For this research, the system and its eventual outputs were not designed with Twitter users in mind; therefore not only does any output have an inherently uncertain impact on those users, the use-case is also divorced from their experience of the system. Accordingly, use of such software, to meet ethical (and moral) requirements, must minimize impacts on users, in terms of likelihood, magnitude and mitigations.

A key principle in determining the ethical implications of the research is the objective, effectively framed by the question: *is the purpose of the research observation or experimentation*? (Grimmelman, 2015:2, Solberg, 2010).

Where the purpose of the research is to manipulate or otherwise influence the actions of research participants or subjects, the burden of responsibility on the researcher is an order of magnitude higher (Kleinsman and Buckley, 2015).

This research has been designed to 'simply' observe, partly in reaction to the type of response received when reaching out to potential participants under the original design. In addition, observation-based research maintains the autonomy of research subjects, and allows the research to continue in the pursuit of an ostensible public good: understanding the realities of policy and public discourse 'better', makes 'better' policy.

In particular, this approach allows a type of digital ethnography at a large scale. As a near real-time platform, discussions and reactions to events arguably happen in a more organic manner than would be possible in, say, a chat room or digital forum, where responses and conversational threads are often time-delayed and asynchronous as people log on and off. By contrast, Twitter uses 'push notifications', messages sent to the user's device (which are turned on by default) to notify them of replies, mentions or retweets. For the site, this helps to keep its users engaged, and from a discourse perspective, this engagement means a more real-time conversation or discourse.

Code Repository

Although samples of the most relevant sections of the codebase, including analysis and data management tools, have been included in the Appendices, the full codebase is many thousands of lines long. A full, structured version of the codebase (with API keys removed) is available at https://github.com/iamwithnail/phd. This also includes 'commits' – a record of the incremental process that took place to develop, test and run this code in fieldwork.

Implementation

A ‘harvester’ programme was created in the programming language Python. The first iterations of this were set up to call the search API and retrieve results on a search-by-search basis, as a proof of concept that the data desired was available from the platform. The initial iteration of the software was set to immediately save a tweet to the database upon receipt. This was satisfactory when tweets were being received at the rate of around 1 per second. As the range of keywords and geographical locations was expanded, the program was unable to keep up, and was disconnected by Twitter; as the program reconnected and was subsequently disconnected, it was then temporarily suspended (‘rate-banned’) by Twitter under their ‘*420 – Enhance Your Calm*’ HTTP response[25], as too many connection attempts had been made.

Anticipating much higher tweet rates, albeit temporarily, another solution was required. In this case, this was to implement a queuing system: tweets were read from the Twitter stream into an in-memory database built on Redis. Tweets were received, a ‘job’ to read them into the database was placed in a queue in memory, and ‘worker’ programmes on the server copied them into the enduring on-disk database for long-term storage. ‘Workers’ are standalone programmes that process information as received – a typical web server will have between one and six ‘web workers’; in this instance, these were queue workers generated using the ‘django-rq’ library, which step through and execute the queued jobs. Operating solely from RAM, this had a speed benefit of several orders of magnitude, and no rate-related disconnects were received. At peak intake, the program was receiving in excess of 30 tweets per second, regardless of how many workers were deployed, and had a backlog queue of 3.2m tweets.

As the database grew, subsequent performance issues became apparent in terms of inserting new records into the database. A bottleneck was caused by

25

insufficient database indexing, and this resulted in a maximum of 1.6 records per second being saved to the database. With a 3.2m record backlog, this would have taken around 3 weeks to read the existing backlog in, without taking into account new tweets, so processing was stopped, and the database was indexed on the 'id_str' field. This is a unique identification field (termed 'snowflake' by Twitter[26]) provided by Twitter for tweets, allowing single tweets to be easily identified and retrieved – this is the number shown in the address bar of a web browser when a single tweet is being viewed. However, this is supplied from Twitter as both a string and as a number, because the API is expected to work with all programming languages, and *'some programming languages such as Javascript cannot support numbers with > 53-bits'.* The original implementation was wrongly configured to use the string version, as Python is capable of handling integers up to 64-bits. String insertion and retrieval is significantly slower than doing so on a natural number, and so this field required indexing to allow research to carry on.

A database index is:

> *"…any data structure that takes the value of one or more fields and finds the records with that value 'quickly'. In particular, an index lets us find a record without having to look at more than a small fraction of all possible records."* (Molina, Ullman and Widom, 2009:619)

A simple (and automatically generated) B-Tree index was added to the database str_id field. A straightforward explanation is that Binary Trees or B-Trees are data structures that imply which 'half' of a database a result is in, minimising the time required to insert or retrieve a record to $\log_2 N$. On a million-record database, this means that the location can be identified in less than 20 searches; previously, to insert records into the research database, the system may have been required to search across the entire table of a million records for each insertion.

26

Stopping the workers (which does not affect queued data, but simply suspends processing it), building the database indices and restarting the process saw this improve, again by an order of magnitude, to around 19 per second per worker. Applying 20 workers managed the backlog queue in around 2 hours; during off-peak monitoring, the program left 2 workers running to read tweets as they were received, and this generally ran with a peak backlog of around 500 tweets. These modifications were successful in gathering the data required for analysis, with the system recording 7,016,114 tweets over the two distinct periods of operation during around protests: the June 20th anti-austerity protests, and the November 5th Million Mask March in 2015.

Summary of Fieldwork Approach

The implementation of a computer program that could operate for long periods of time gathering data of interest to the researcher was comparatively simple. In the period since the fieldwork was carried out, a number of additional libraries and services have been made available that allow access to such data without recourse to programmatic methods, although this would still be the preferred option in terms of low-level access to the data. Commercial services such as Hubspot, Hootsuite, AgoraPulse and Keyhole allow individuals and brands to monitor hashtags, keywords and other trends on social media, particularly Twitter. These provide an excellent entry point into understanding the data that is available to the researcher, which can then be combined with other approaches and software to enable the capture and retention of this data for further analysis after the fact.

While a significant volume of data was gathered through the approach laid out in this chapter, the next section demonstrates the need to check the gathered data for validity at or around the time of capture. Streams and searches should be prepared carefully to ensure that data noise is minimised, since the reliance on filtering and cutting down extraneous data at a later time can be problematic, as explored in Chapter Eight.

Chapter 8:

Analysis of Fieldwork

Introduction

This chapter sets out the approaches that were used to analyse the data gathered as a result of the methods in Chapter Six. It highlights a number of different strategies that were utilised to attempt to gain insight from the data, as well as explicating difficulties faced in doing so. As outlined in the methodology chapter, the approach had been to leave the 'harvester' running for a number of periods over the summer of 2015, monitoring specific keywords around surveillance. As protests approached, or were announced, the keywords and area monitoring were adjusted to take advantage of these new areas of interest, with the hypothesis that there would be an element of coalescence around debate during these events.

In light of the difficulties that were faced during the analysis due to oversampling and poor survey design, the latter half of the chapter lays out the lessons learned through working with Twitter as a data source. It also examines in some depth proposals for future work, and discusses the practical implications of the limitations of the research, and how they can be avoided and mitigated in future work.

Response and Relevance Rate

The number of tweets over this time period, per hour, is shown here from June 6th to November 6th 2015. The gap indicates the time between the two protests where data was not being gathered, but gives an overall indication of the relative velocities of tweets during and around the two protests.

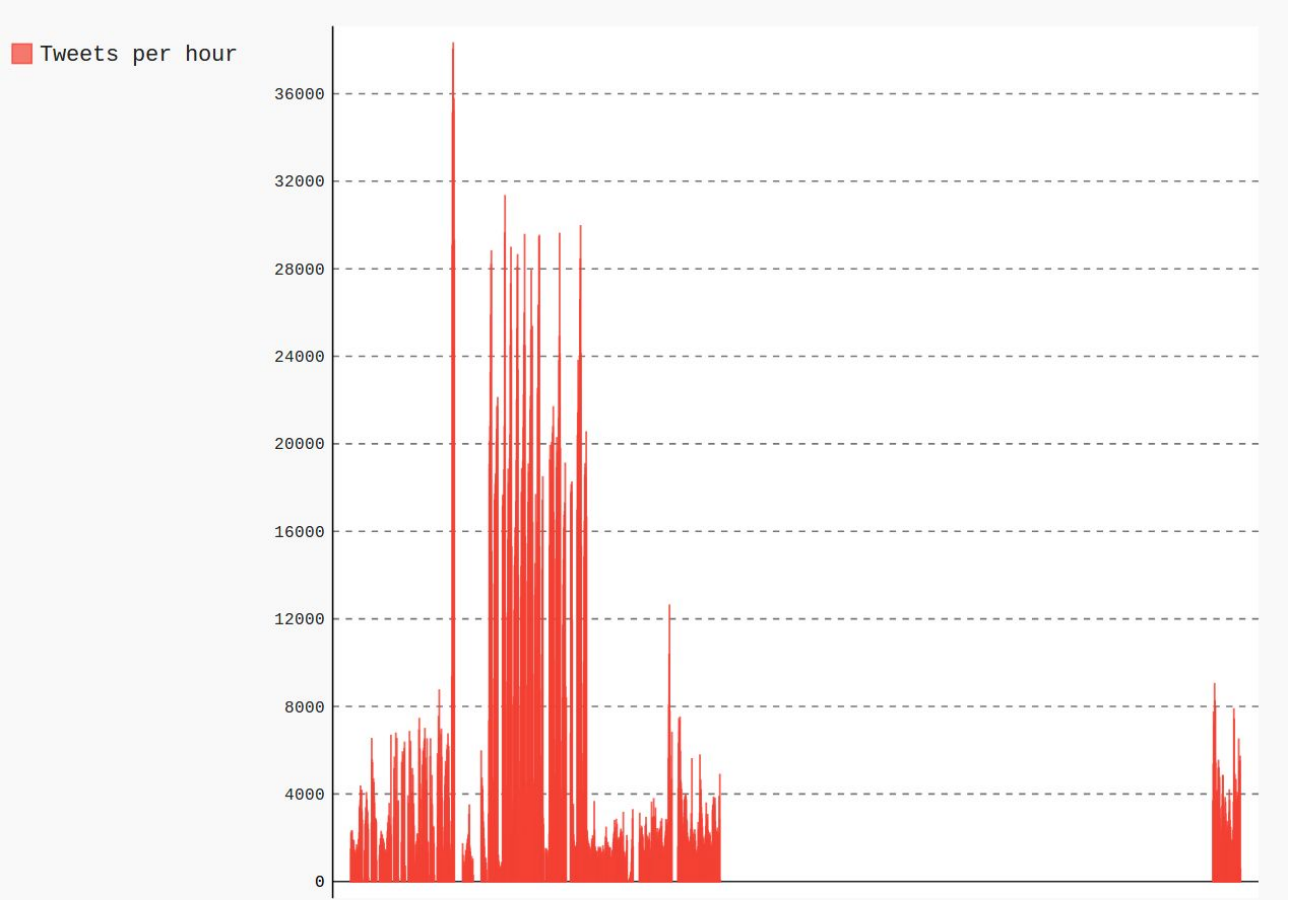


Figure SEQ Figure * ARABIC 1 - Tweets Per Hour – June to November 2015

However, as indicated below, a number of keywords provided large numbers of international tweets that constituted unhelpful noise in our overall data set. As part of that work, it is possible to exclude those terms. The most straightforward way to do this is to apply the filter on the initial 'whole database' dataset, and then select the time periods or other features of interest from this smaller, revised dataset. However, this is slow to run in comparison to a straightforward count by time, and runs into (fatal) memory capacity problems, which are detailed below in the 'Performance Issues Affecting Analysis' section.

The alternative approach to take is to analyse each time slice as its own database query, and apply the exclusions filter against that. While this avoids the issues of overall memory space, because the whole data set is not being loaded into memory, it is even slower to run. While the 'tweets_by_hour'

function ran in around 15 seconds over the entire dataset, taking a database slice and then applying a series of exclusions based on 'OR' statements took dozens of hours to run over the same time period, processing, on average, one hour's tweets every three or four minutes. With 3,777 hour-long periods available in the dataset, this was similarly untenable.

As with the tweet backlog noted in Chapter 6, the immediate solution to this problem was to break up these jobs into functions that could be placed on a queue. The analysis processing software was then cloned to a number of machines available to the researcher, which were able to interface with the queue and process the jobs. Results from each of these jobs was written to a central store using a key-value pair, with the 'created at' time of the first tweet in each slice used as the key. Whilst still slow, this has the added advantage that work can be paused and recommenced, with failed jobs being visible and available for re-queuing and reprocessing in the event of failure. Jobs which do not complete successfully (that is to say, jobs which exit without a success code) are placed in a 'failed jobs' queue. The researcher was able to evaluate the reasons for failure (such as temporary loss of network or database access), and requeue the jobs as necessary. Filtering the overall dataset in this way reduced the high-level dataset from 7,016,114 to 6,647,999 tweets, a reduction of only 5.2%.

Tweet Data Structure

Tweets are, perhaps surprisingly, complicated data structures. The following table lays out the data structure of a tweet as stored in the database:

Table 1 - Tweet Data Structure

Tweet field	Notes
truncated	Boolean indicating whether the tweet has been shortened to meet

	length requirements, because it includes, e.g. a URL or other shortened format. Not visible to users.
text	The recognisable body of the tweet: ‘This is my first Tweet, hello world!’. Visible to users.
in_reply_to_status_id	Adopted for threading of tweets, allows retrieval of tweets that form part of a ‘replied to’ conversation. Indirectly visible to users as of 2016; clicking into a tweet in mobile applications or the web UI will show the tweets corresponding to this id.
id	Internal research system ID number.
favorite_count	Integer representing the number of times the tweet has been ‘liked’ or ‘favourited’. Visible to users.
source	Text field indicating the type of application tweet sent from – Web, Twitter Mobile App, Tweetdeck, etc. Visible to users.

_json	JSON[27] representation of all tweet information in a single field, predominantly used by mobile applications. Not visible to users.
coordinates	The latitude/longitude of the tweet sent if geo-location is turned on, or corresponding to Place if set.
in_reply_to_screen_name	Early implementation of threading, shows the screen name of the user that that this tweet was in reply to, but does not link to the tweet directly.
id_str	Twitter ‘snowflake’ unique identification number of tweet, represented as a string to allow programming languages without large signed integers to correctly handle the ID. Visible to users as part of the url of a tweet.
retweet_count	Integer count of the number of times a tweet has been retweeted. Visible to users.

27

favorited	Boolean indicating whether the *accessing* user has favourited this tweet. Visible to users.
retweeted_status	Boolean indicating whether the *accessing* user has retweeted this tweet. Visible to users.
user	An expanded representation of the tweeting user's data, including description, location, followers and follow counts. Visible to users in summary and as link to user profiles.
geo	Geographic data in simple format, subset of Place.
in_reply_to_user_id_str	Early implementation of threading, shows the id of the user that that this tweet was in reply to, as a string, but does not link to the tweet directly.
possibly_sensitive	Boolean indicating whether the tweet contains media (images, audio or video) that may possibly be sensitive.
lang	ISO Code Representation of tweet language (set by tweeting user,

	and used to offer in-app translation where different from the *viewing* user's set language.) Visible to users.
created_at	Date time object, created from the string timestamp provided by API. Visible to users.
in_reply_to_status_id_str	String representation of the ID of a replied-to status. Adopted for threading of tweets, allows retrieval of tweets that form part of a 'replied to' conversation. Indirectly visible to users as of 2016; clicking into a tweet in mobile applications or the web UI will show the tweets corresponding to this id.
place	Long format place detail, such as city, location type (point, polygon, bounding box).
retweeted	Boolean indicating whether this tweet has been retweeted.
author_id	Twitter ID number corresponding to the user who created the tweet. Not directly visible to users.

Age Distribution of Sample

In May, July and November 2015, the age demographics of Twitter were distributed as follows[28]:

	April 2015	July 2015	November 2015
55+	12%	7%	10%
45-54	13%	16%	15%
35-44	20%	24%	19%
25-34	23%	24%	17%
15-24	33%	30%	29%

Table 2 - Age distributions of Twitter Users - Spring to Autumn 2015

This compares unfavourably to the overall population distribution of the UK as highlighted in the 2011 census (Office for National Statistics, 2011). Younger age groups are disproportionately overrepresented, while older groups are less present.

55+	28%
45-54	14%
35-44	14%
25-34	13%

28

15-24	13%

Unlike other social media platforms like Facebook, and various predecessors like MySpace and Bebo, Twitter does not explicitly display age in a user's profile (in part due to the platform's public-by-default approach). This makes it extremely difficult to adjust the sample and results to understand how this specific sample relates both to the population at large, and against the UK Twitter population more generally. Morris (2015) notes that fewer than 0.5% of Twitter users note their age on their profile.

Although it is possible to use a machine learning approach to classify users into broad age groups (Morgan-Lopez, et. al., 2017), this is a significant challenge in its own right, and is not an approach that has been adopted by this research. Morris further notes a key difficulty with this type of biased sample, in relation to the Scottish Independence Referendum in 2014, that the impact of the 'silent majority' can be significantly underrepresented in modelling and polls. This potential impact has been borne in mind when conducting the rest of the analysis.

Political Alignment and Affiliation

With 1,309,804 users' details collected, any mention of a political alignment was extracted from the user's biography (*'description'*), to ascertain if any judgement could be made as to the overall make-up of the sample, potentially for comparison with the outputs of the BES 2015, which would have provided a contemporary source for comparative analysis. However, this search only yielded 9,160 users, or 0.69% of the total user base, many of whom were based in the United States (the difficulty of reliably separating Americans who self-identify as 'Conservative' against those in the UK who may identify with, vote for or are members of the Conservative Party). In turn, we can hypothesise that, given the broad relationship between age and voting

patterns, older voters are more likely to have voted or support Conservative policies (YouGov, 2015), and that there will be a similar underrepresentation of these views on the platform.[29] Again, although it may be possible to identify, and so give additional prominence to these users (Tumasjan, et al, 2010) in the analysis, this is outwith the scope of this research.

Distribution by Users

The distribution of tweets amongst users was also strongly skewed, with the top 1% most prolific users providing 2,328,907 of the overall 7,016,114 tweets, or 32.7% of the total. A significant number of these were non-UK based. In part, this issue is driven by a decision to use separate streams of keyword and geographical bounds. Using keywords, rather than hashtags, to search for a term gathers a lot of 'noise' in the data sample, where individuals have mentioned a keyword in passing, but did not explicitly tag it as relevant. Bruns and Stieglitz (2014:74) note that the while a "*..keywords dataset… constitutes a cross-section through the Twitter activities of users who are largely likely to be unaware of one another… hashtags inherently provide at least the potential for such awareness… Keyword datasets necessarily contain fragments of wider conversations… and their metrics must be understood from that perspective.*" In part, the difficulties and limitations noted later in the chapter arise from this issue. The dataset was gathered without flagging which search it resulted from, which renders analysis of the dataset in totality far more difficult.

Additionally, although the skew in favour of more prolific users was picked up in the initial analyses, the analysis itself failed to make any substantive use of that. As noted from page 186, there were a number of analytical approaches that could have been used to capitalise on this feature, rather than treating it as an inconvenience of data.

29

Graph Analysis

One option for analysis for Twitter data is to place tweets or users into a Graph network tool, such as a NodeXL or Gephi, programs that allow user or tweet datasets to be gathered in Excel sheets and imported into the network tool. Graph analysis provides that networks are made up of nodes and edges. In a social network like Twitter, nodes could be users, while the edges are the relationships of 'following'. The intention was to utilise such a system batch to identify the shape of discourse by analysing tweets and relationships between tweets: which tweets were replied to, which ones were retweeted? Which tweets elicited strongly negative or positive responses? This would allow the creation of snapshots of the discourse at various times, showing the (literal) shape, indicating who and what phrases are most central to the debate at any given time. Understanding the relative levels of interactions for specific tweets would also allow the possibility of weighting that text in other analyses, to account for its importance in the network.

As the tweet data had already been collected, it would be a comparatively simple matter to show these networks using, NetworkX, a python library that allows the creation of edges between objects. The Twitter API offers two fields that are relevant to creating this analysis: `in_reply_to_user_id_str` and `in_reply_to_status_id`. However, although this is available at the time of submission (late 2017) and was offered as a field at the time of data gathering (summer to Autumn 2015), this feature was not fully implemented at the time. Replies (called '@ replies', where a user places the @ symbol in front of another users handle to notify them of a reply) have been a Twitter feature since 2007 (Twitter, 2007), but were not publicly available to users of the API in 2015. None of the 7m+ tweets that were gathered through this research had any entries in the 'reply to status id' field, meaning it was not possible to trace the reply history through these specific tweets.

The ‘In reply to user’ field was populated, and was usable, but this only provides a basic indication of who is talking to whom; without the link to the specific tweet ID, comparing and gauging the relative levels of response is almost impossible. There were 1,979,393 tweets that had an ‘in reply to user’ field set, or 28.2% of the overall dataset – this would be an extremely useful feature to implement in future research. It is worth noting that Twitter introduced the concept of ‘threading’[30] into the main web browser and mobile application in March 2014 (Twitter, 2014), but it had not yet reached the free tiers of the API by summer of 2015, although it was available for in-app development under the paid tiers.

Sentiment Analysis

To circumvent the analysis issues caused by the unavailability of conversation or ‘thread’ data, the research attempts to classify the data in terms of sentiment analysis. Sentiment analysis is a computing technique that attempts to classify whether a textual expression is positive (favourable) or negative (unfavourable) towards the subject (Nasukawa and Yi, 2003).

Sentiment analysis takes the form of the researcher manually assessing text data and assigning it a positive or negative rating, to create a training set or classifier. With a training set created, the researcher can then parse further text through the classifier to assign new text data a rating. This can typically be in the range of 75-95% accurate, depending on the text data (ibid: 76). For this research, this was of interest as it would allow not just the shape and general content of the discursive assemblages to be assessed, but also the relative *direction* and *strength* of feeling over time; it would also allow, using Latent Semantic Analysis as described below, the similarity of the terms used on opposing sides of the discussion to be analysed.

A recursive trainer was created, which took a random set of 5,000 tweets from the database, and presented them sequentially to the researcher for

30

classification. These were then to be classified as positive, negative, neutral, or irrelevant. If a tweet was marked as irrelevant, then the tweet was broken down ('tokenised') into keywords, with common stopwords (such as 'the', 'a' - a list of stopwords is provided in Appendix Three) removed. These remaining keywords were then filtered from the remaining tweet set, and the analysis continued.

It quickly became apparent, however, that the dataset was too noisy to allow this method to work; due to the high level of irrelevant tweets, the exclusion filters for keywords quickly became too large to operate with any kind of efficiency, and after the first ten to fifteen tweets had been analysed and classified, the database queries slowed to an effective halt, rendering further analysis impossible. However, this is an interesting and potentially extremely valuable technique if it can be usefully cross-referenced with other techniques, and should be re-evaluated for further work. The analysis code for this type of analysis and data management is available at Appendix Four.

Latent Semantic Analysis

Latent semantic analysis (LSA) is a technique for parsing and analysing the similarity of words and sentences in a corpus of text (Dumais, et. al., 1988, Deerwester et. al., 1991, Landauer, Foltz and Laham, 1998). It has been shown to show significantly deeper relationships between the text elements than simple co-occurrence counts or correlations in word presence and usage, and also significantly reduces the dimensionality of information retrieval problems – in other words, it can recognise multiple terms for the same object (synonymy) and a single term with multiple meanings depending on surrounding context (polysemy) (Dumais, 2004). It is particularly useful in this kind of social media analysis, as it can evaluate a text without additional context, Landauer et al arguing that:

> *'...the representations of passages that LSA forms can be interpreted as abstractions of 'episodes'… its representation of*

> *words, in turn, is intertwined with and mutually interdependent with its knowledge of episodes… while LSA's potential knowledge is surely imperfect, we believe it can offer a close enough approximation to people's knowledge to underwrite theories and tests of theories of cognition."* (Landauer, Foltz and Laham, 1998:5)

In the LSA process, the text or texts to be compared are represented as a matrix, with each distinct word in a row, with corresponding columns for the textual passage. Each matrix cell represents the frequency of occurrence of each word in the corresponding context. In turn each of these values is subject to singular value decomposition. There are well-established tools that allow non-mathematicians to use such methods. This research utilises the python library developed by KernelMachine at the University of Washington (KernelMachine, 2015) to do just this. A fuller introductory explanation of the process by which LSA and its constituent mathematics work is available at Figures 2-5 in Landauer and Dumais (1997))

In practical terms, the output of analysing two textual fragments against each other using this technique is a number ranging from 0-1 to indicate semantic similarity.

The rationale for applying this technique within this study is that seeing the 'shape' and 'feel' of the discourse over time will allow some insight into the manner of change. From the perspective of considering the discursive or expressive elements of an assemblage, LSA similarity can be hypothesised as an analogue for territorialisation of debate. As the semantic similarity of the debate increases, particularly over the average, it is reasonable to assume that more people are repeating the same or similar things: the debate is becoming more territorialised and fixed, and potentially heading towards a point of crystallisation.

Similarly, there are instances or moments where the level of semantic similarity suddenly changes, and this can be hypothesised as a line of flight

has been introduced that is deterritorialising the discourse. These hypotheses are re-examined in light of the two case studies where this analysis has been carried out below.

Conversely, a lower level of semantic similarity may indicate that the debate is shifting, or indeed that there is little cohesive discussion. This latter point is one of the areas where examining the relative shape of the discourse would be useful – examining whether the discursive fragments are dispersed in small clusters, are simply isolated, or are discussing a range of things to do with the same topic, but in very different ways – by allowing us to understand the nature of that low level of cohesion, and in particular, trace it back to a previous moment of high cohesion.

Data Noise

As may be expected in a large data set such as this, there was and remains a large degree of noise, which is defined here as 'data collected which is not necessarily or directly relevant to the issue being examined'. As outlined in the previous chapter, this research deliberately chose to oversample the available data, due to the time-limited nature that tweets and data are easily available.

One of the keywords that had been used to capture data was 'arrest', on the straightforward assumption that this would capture any significant action by the police. However, this specific term proved particularly problematic, even over the relatively short periods of capture and analysis. Over the period, there was significant news from India around the Modi government, with several prominent officials in the BJP being arrested. As Twitter is a significant news source for many in South-East Asia, this was covered by numerous news outlets, and re-tweeted by many of their followers. In addition, this unfocused keyword also collected tweets from various news outlets in the US, with local affiliates of large news stations (Fox, ABC, CBS) either tweeting

the same headlines as they broke, or retweeting content from other affiliates and central accounts.

In addition to these sources of noise as a result of the over capture principle, there was also some degree of commercial spam present in the collected sample. This occurs where marketers place product links in a collection of popular and trending hashtags, in order to make the links visible and receive clicks from unwary users who have been browsing a particular hashtag.

There is a qualitative assessment to be made: do these tweets constitute valid discourse and discussion, and so should be included in the analysis datasets, or are they unnecessary noise, which should be excluded? There is a distinction to be made between outright (and obvious) commercial spam that piggybacks on popular hashtags, and that of 'social spam'. This is spam which is often generated at an industrial rate by farms of automated accounts which manipulate existing (and therefore more trusted) accounts into posting content and links on their behalf, either through social engineering or outright hacking (Chu, Widjaja and Wang, 2012). In many cases, this may be less clearly spam, and may pose as a user.

Chu *et.al.* view spam as a polluting feature of online communities, which seems to be valid; it detracts from the purpose of the community, even one like Twitter which sees multiplicities of conversations and discourses operating both simultaneously and asynchronously, as areas of activity move round the globe over the course of the day. Where spam keywords could be clearly and unambiguously identified, they were removed from the dataset, as were tweets such as those described above, which are interesting but irrelevant to the subject both in terms of geography and content.

To do so, sets of tweets were analysed by hand, displaying them on the screen in batches of 100. Anything that was clearly from outside the UK or unrelated to the protest subject matter was highlighted, and key phrases or terms were copied to generate a list of exclusion phrases and keywords. From

this list, a database filter was created on an ‘OR’ basis. Any tweet that matched any of the keywords or phrases from the list was excluded.

As this is a crucial filtering operation, which is used repeatedly, the exclusion list is shown here, with the code being available in Appendix 2.

```
['cardiac arrest', 'CARD/RESP ARREST',
'#fit','#androidgames', '#iphonegames',
'http://t.co/1aG5q2YMlw',
'#ipadgames', 'Good Gut Bacteria', '#fatloss',
'#instadeily',
'asaram bapu ji', 'Maher Arar', 'Michael Barrymore',
'Delhi govt raided JIMS Jagran Institute of management
studies in Rithala',
'BJP', "Uganda's ex PM defiant after arrest",
'Tomar', 'Nebraska Murder', ' shooting of incoming BYU
recruit','#aap', 'Nigerian Troops Arrest Jos',
'#GameInsight','Nigeria Arrest Former Chief Security
Officer', '#healthylife','#mystyle', '#ootd',
'#instagramers',
'Buhari reportedly planning to arrest', 'kpk', 'EFCC
confirms arrest of former governors Ohakim', 'Amama
Mbabazi', 'JahangirKTareen', 'Godwin Obua', 'Manoj
Kumar', 'Muhammadu Buhari', 'Federal Marshals Arrest
12-Year-Old, and Accomplices', 'Federal Agents Arrest
Baby-Faced Boy', 'Taeba Darwish', '#jungkook', '#jin',
'#suga', 'Peoples Democratic Party',  'Xiaomi Mi5',
'Moses Kuria', 'Police arrest masquerade in Enugu for
stabbing priest',  'Layla Al Qaseer', 'Raihana Mosawi',
'Jalela sayed ameen',  'Nigerian Troops Arrest Mastermind
Of Jos', 'in Karachi operation' '#FitnessFriday',
'#girlswholift', '#deadlift', '#fitfam', '#Palestinian',
'#ModiNotWelcome']
```

It is interesting to consider the experience of Twitter users in relation to data noise. The bulk of the day-to-day use of the site for users is generally through the feed, where users see the tweets of accounts that they have chosen to

follow, originally in chronological order. This is ostensibly a self curated and opt-in feature, defined by the users themselves, and so comparatively free of 'noise', which can be defined from a user experience point of view as 'Tweets that I did not expect to see'. Twitter have made a number of changes in recent years to how this feed is curated from a system point of view, defaulting to 'Top Tweets' rather than chronological. These 'Top Tweets' are curated by an opaque system on behalf of the user on the basis of their use of the site and the 'signals' this generates. [31]

While there is a clear difference between what is shown to an individual user, and the data that is gathered from a programmatic and systematic search, the delta between these two sets of data is unclear, which makes it difficult to draw comparisons in terms of noise. The presence of data noise seems to be apparent in the number of measures that Twitter has taken to provide more relevant and engaging content to users, including the Top Tweets default feed, 'In Case You Missed It' summaries when the user logs on after an absence, as well as adding *other popular content* to the user's feed. Even while scrolling what appears to be their own feed or through a hashtag search, users are now being shown a heavily curated and managed feed of the site. The comparison between a user feed and what is gathered from accessing site content through the API would be worthy of further investigation, simply to understand *what* the difference is, and to allow some insight into how this could affect public discourse around topics with varying degrees of popularity or engagement.

Geographical Use

The Twitter stream was captured along two dimensions – a set of keywords deemed pertinent to the research subject, and along geographic lines. The intention was that this would allow analysis of the changing nature of the discourse to be calibrated against the overall flow of online discussion. However, this was an unsatisfactory decision, and introduced both noise and

31

ambiguity into the data set. Upon collection, there was no consideration given to setting a flag or marker on individual tweets to track its source of origin – effectively, to mark out 'what had caused this to be selected for harvesting?' This means that the overall data set is an amalgamation of geographically selected tweets from around London, and keyword tweets from around the world. The latter of these is doubly problematic, for the reasons outlined above.

Twitter provides three sets of location metadata in tweet information: coordinates, place, and geo-data. Coordinates are straightforward and represent either a specific point, or are two sets of coordinates that provide the southwest and northeast coordinates of a bounding box around the user's location. 'Geo' information provides details about the type of information that is held, indicating whether it is a bounding box, coordinates, or other type of information. Place, meanwhile, replicates some of the same functionality as the other boxes – descriptor of location type, coordinates, and further details about the area, but is set by the user, rather than through the phone or device's location settings. This means that, in some instances, the user data is wilfully (if not necessarily maliciously) incorrect, or changed to provide satirical or humorous content: as an example, there were over 100 tweets whose location was set as 'Hell'.

There were 611,114 tweets that had coordinates data embedded in their metadata, or 8.7% of the total gathered; cross-checking this against tweets which had geo-data enabled returned the same set of tweets. In contrast, there were 4,442,021 tweets that contained 'place' data, which as indicated is set by the user. Being user-managed, the location refers to time the tweet information was received from the API, as Twitter gathers the data internally by joining to the user's profile information. This is less of an issue for the continuous gathering of data through, for example, the Twitter Streaming API, as this fetches the information for each tweet – it is live and contemporary. However, revisiting a tweet later, for example to retrieve additional information that may have been missed, or to check subsequent levels of retweets, the

location data from ‘place’ will refer to the user *as currently set*, not at the time the tweet was created. Geo and coordinates data are, however, effectively location stamps for that particular point in time.

Limitations

Due to the bounding box, the particular density of Twitter users within London is notable (Longley *et al*, 2015), which creates an inherent bias in the data of the study, which has been specifically targeted at users within the London ‘bounding box’, a square geographic centred over the M25; issues that are of high importance within London may not be reflective of those in other urban conurbations throughout the UK, or the UK as a whole. London contains a critical mass of social media users on this particular platform, which may lead to more in-depth, wide-ranging, or sustained conversations than in areas with a lower level of platform uptake.

Although Twitter has a high level of uptake in the UK than globally, it is as indicated earlier skewed towards younger demographics. This study does not take account of other content-based social media such as Facebook or Tumblr, which would also have usable textual information that arguably forms part of the overall discourse. However, these methods are more private, which raises further ethical questions as outlined in Chapter Six.

A suitable mitigation to the noise and geography issues could have been to filter tweets as they were received from the stream, and either exclude them based on geographical metadata, or on the basis that they did not include the search terms, if they were in the appropriate area. However, with the performance of the tweet harvester system already under strain, it may have been difficult to actively manage this filtering at the point of collection. These issues are explored in the final chapter.

Case Study One – Anti Austerity March, June 20th, 2015

During the summer of 2015, there was a large-scale protest arranged for Saturday, June 20th, to protest the Conservative government's ongoing austerity measures, which had been in place since the 2010 election, and were renewed after the narrow Conservative victory in the 2015 May election. Initially, this study attempted to analyse tweets from 00:00:00 18th of June to 23:59:59 on 22nd June (2 days either side of the End Austerity Now protests), yielding around half a million tweets. Initial examination of these for the most common keywords (with search keywords removed from the corpus).

Table 2 - Most Common Words

People	19,340
March	15,567
Austerity	14,085
Surveillance	13,806
London	13,134
Arrest	11,464
Demo	10,370
Day	9,926
Love	9.775
Demonstration	9,005
Time	7,566
Thousands	7,431

Police	6,436
Join	6,023
@pplsassembly	5,894
Look	5,484
Thank	5,372
Please	5,269

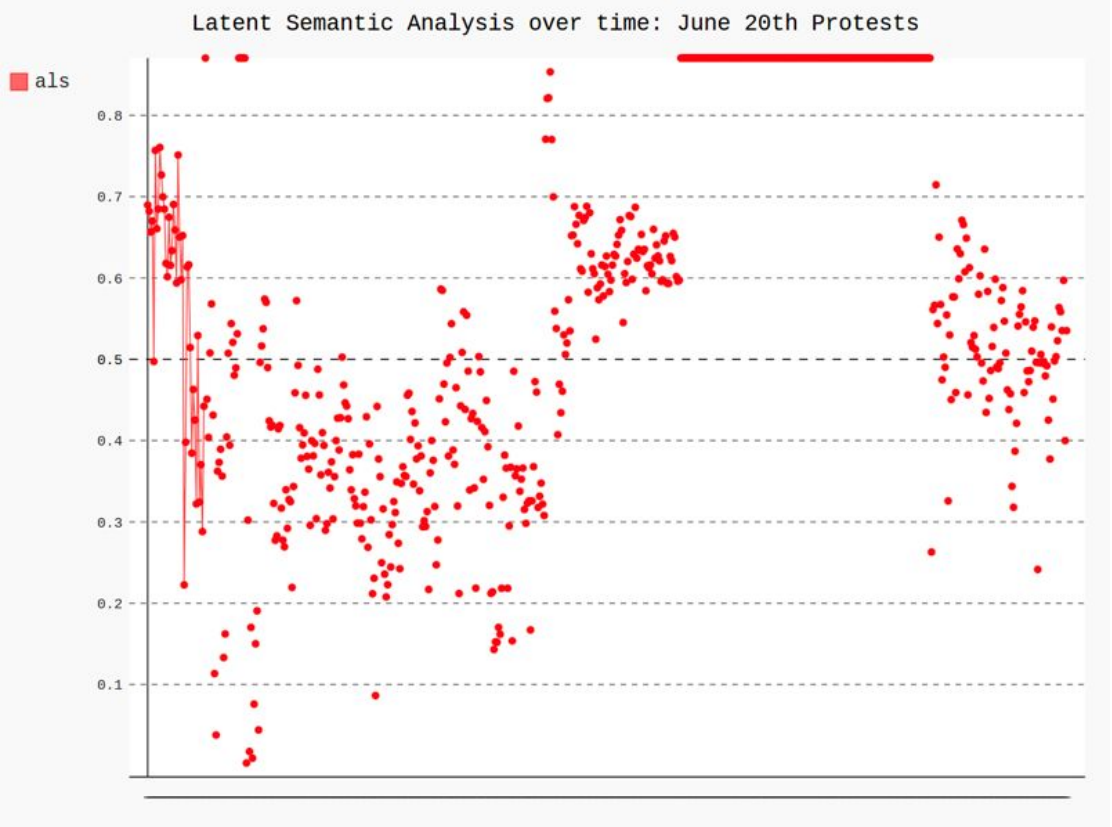
However, such a broad set of search terms, as indicated above, proved unhelpful in trying to create a corpus of surveillance and protest-related tweets. The 10-minute slices reflected the overall data around them too closely, as there was no filter to specifically compare protest-related tweets with the general discourse.

The analysis was therefore refocused to pull matching keywords, which were 'stemmed', so that, for example 'demonstrat' would match 'demonstrate', 'demonstration', and similar partial matches. This function is available in the `analysis.tools` module, under `june_20th_tweets`. The search was widened to draw a more focused comparison based on keywords, but drawn over a longer period, from June 6th to June 22nd, to allow examination of how this topic emerged from the overall topics and discourse occurring online.

This keyword set matched 204,376 tweets, which were divided into ten-minute segments, between the beginning and end. Each segment was reduced to a corpus of words, with the surrounding half-hour (fifteen minutes either side of this slice) of tweets being drawn, cleaned and similarly reduced to a corpus.

On analysis, an issue was discovered with the dataset collection. Due to an unforeseen software error in the harvester program, there was a gap in the data. No tweets had been collected between 17:43 on June 20th and 20:52 on June 21st. Although this is ostensibly after the main protest period, with the march scheduled to run from 12:00 to 15:00 in central London, it is significantly unhelpful in examining how the protest content relates to overall discourse over the period.

Table SEQ Table * ARABIC 3 - Latent Semantic Analysis over time - June 20th 2015



Broadly, the analysis of rolling periods of coverage shows an increase in the semantic similarity of the target content against the broader sample. This is to be expected; the protest was a large-scale event, which had dominated much of the national media in the days running up to the event. There is a surprising cluster of semantically similar tweets on June 6th, the first day of capture; manual examination of tweets from that day seems to indicate that this was

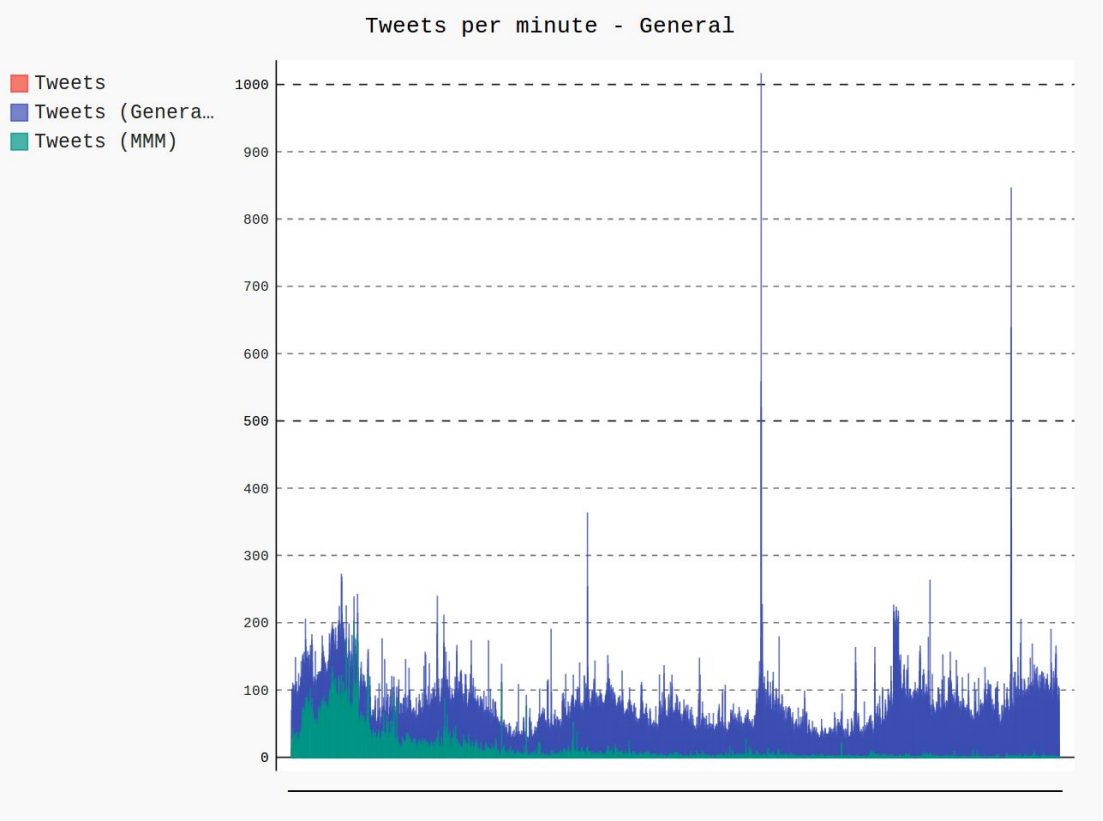
the day the protest was announced and re-publicised, with a large number of people sharing content from the organisers. As both this release date and the protest itself were on a Saturday, these are comparatively 'slower' news days; the assertion should be made that this allows a comparatively niche discourse (of organising the protest) to more effectively dominate the online space, and this is a hypothesis that would be worthy of further work and investigation.

In light of the gap in analysis across the period, the analysis was run over a longer period, covering the week before and the days themselves. This was particularly difficult to achieve from a computational basis, as the datasets were simply too large to hold in memory, and this required more active memory management than is typically the case.

Case Study Two - Million Mask March

Data was selected based on the presence of three key phrases: 'million mask march', 'mmm2015' (to capture the hashtag), and 'millionmaskmarch'. Each search was case insensitive. This search yields 76,186 captured tweets. Tweets started at 12:09:36 UTC on 5th November 2011, and finished at 19:53:48 on 10th November, for a total of 7,664 minutes. Filtering for spam keywords and phrases, as discussed above, did not reduce the total in this instance.

Table 4 - Tweets per Minute – Summary during Million Mask March



The initial analysis attempt was to cycle through each tweet, and compare it to the surrounding body of tweets from the 'general' tweet population for similarity using LSA. (Code shown at Appendix Three). For each tweet, all tweets within a time delta of 15 minutes were selected. The initial tweet and its surrounding counterparts were reduced to two tokenised corpuses, which were then compared using latent semantic analysis.

This approach was limited by the computing time (in the region of 5-10 seconds for each comparison: 4 to 8 days of compute time), and by the practicality of requiring to use sub-second scale times. Across the 7,644-minute-long period of analysis, an average of 9.94 tweets per minute were recorded. These were distributed extremely unevenly across the time period, peaking at around 200 tweets per second, making it particularly difficult to represent the LSA results over time.

Table SEQ Table * ARABIC 5 - Latent Semantic Similarity to surrounding tweet corpuses

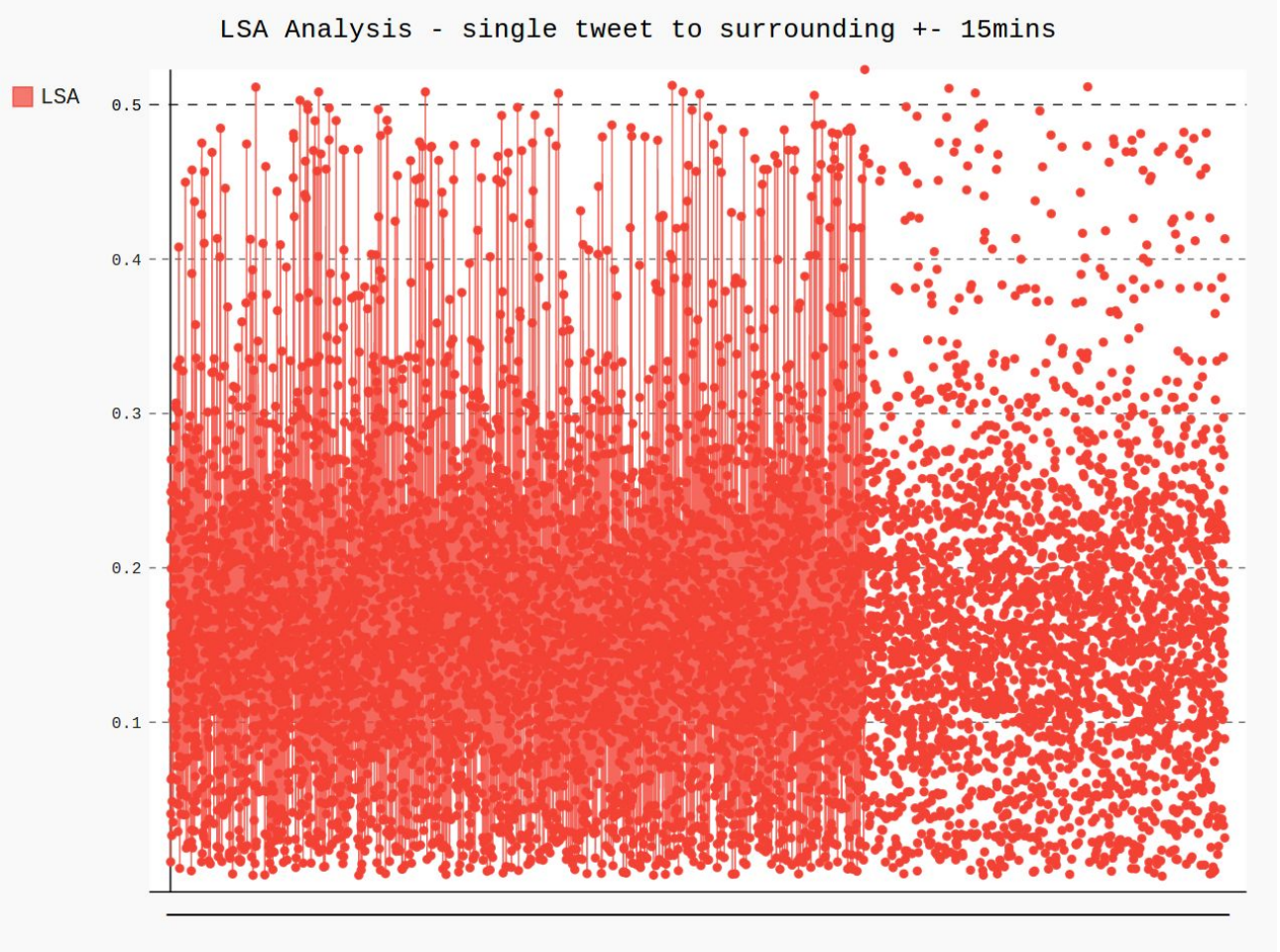
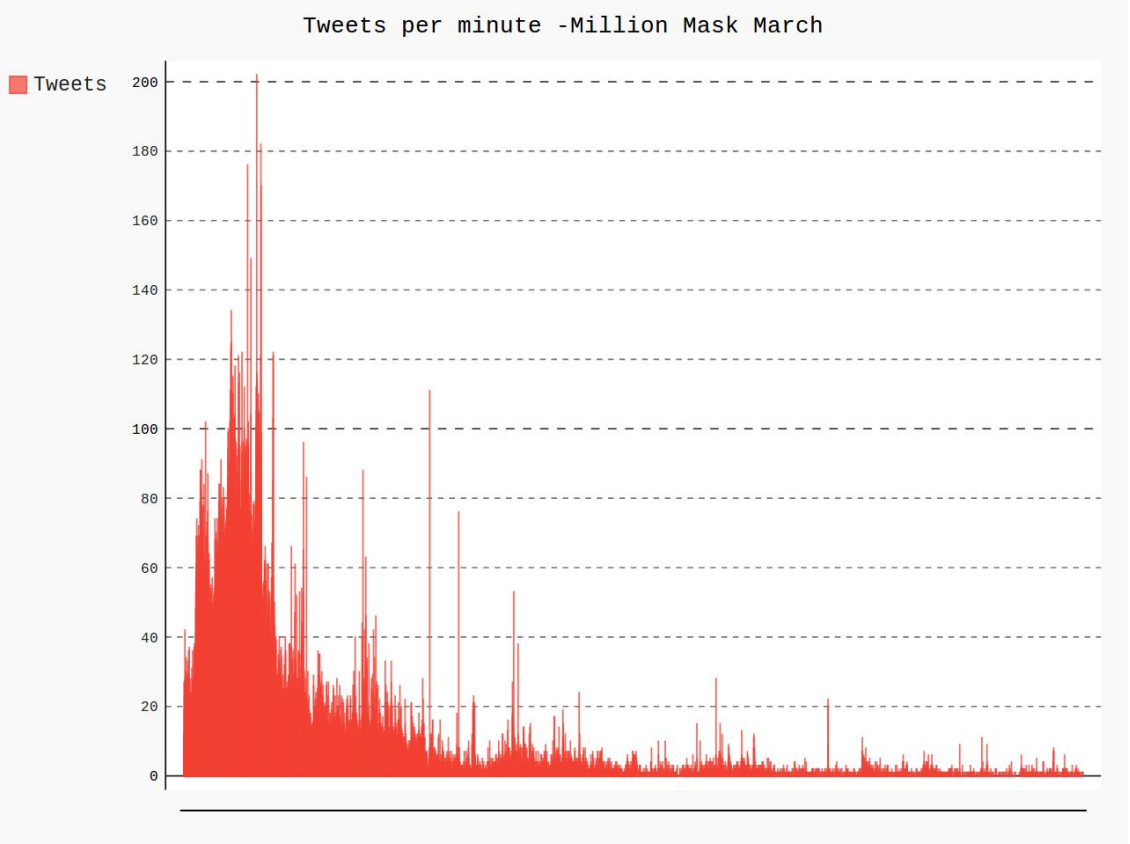
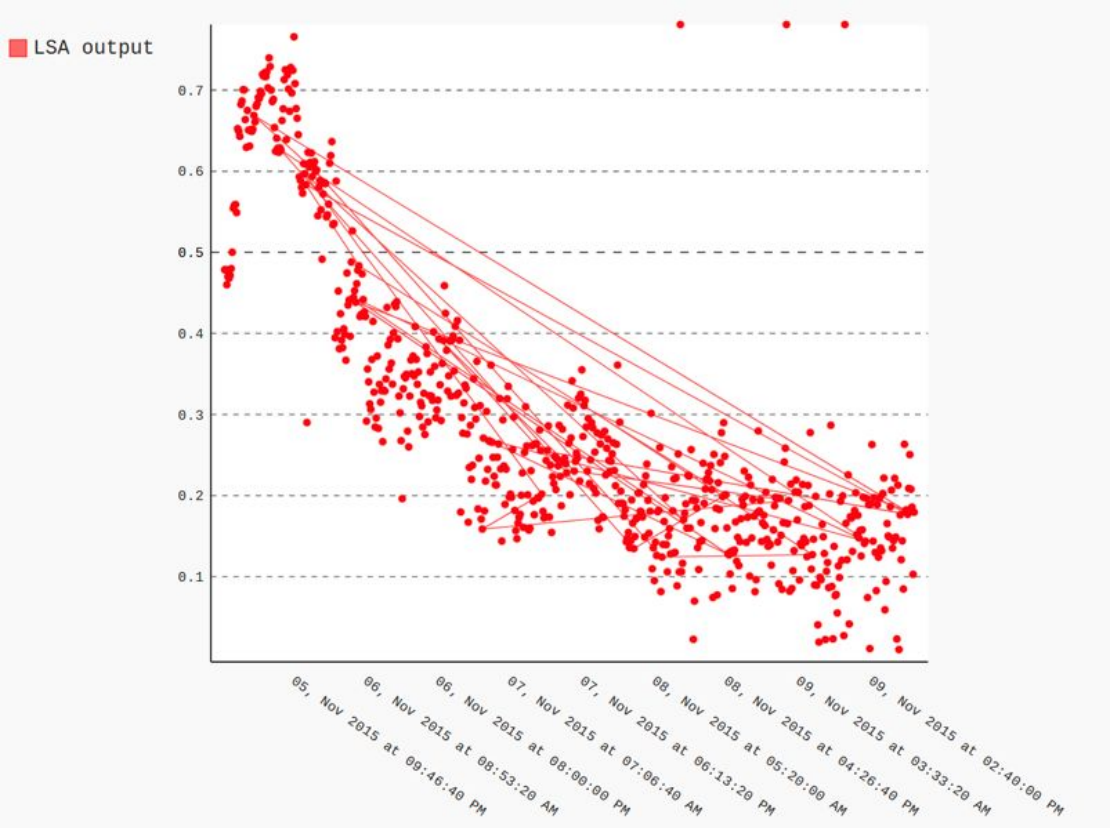


Table 6 - Tweets per minute using #millionmaskmarch



Instead, the set of tweets available for analysis was bucketed into 10-minute intervals, with the surrounding tweets within a time delta of thirty minutes being the comparator.

Table 7 - Latent Semantic Similarity over time - Million Mask March



This shows that the tweets in the run up to (6pm GMT) and during the protest (evening of the 5th of November) have high levels of semantic similarity. To some degree this is to be expected; as with the June protests, it was a large event, occurring in central London, although the Million Mask March was a coordinated and rolling worldwide movement. This is reflected to some extent in the extremely high 0.7-0.8 similarity scoring on the evening of the protest, peaking around 10pm GMT. In contrast to the June protests the scores drop rapidly, receding below 0.4 by 8am GMT the following day, and continuing to fade over the course of the week.

Performance Issues Affecting Analysis

It is worth noting the (significant) on-going performance issues that the study faced in terms of analysis. Because keyword searches are looking across a `CharField` (text string) entry using the regular text lookup method,

performance was low; in the order of 90 seconds to run a keyword search and entries count. This renders some of the more interesting analysis tools difficult to run, as simply getting the information out of the database at any sort of scale is problematic. Leaving programmes to run for long periods of time is an option, but this is vulnerable to bugs and other failures. If a programme or piece of analysis that is set to run for several hours – or overnight – encounters an issue, then some or all of the analysis results can be lost. Although this can be minimised and guarded against by proper error handling, many of the improvements in performance would require significant steps towards high-performance computing and distributed systems to adequately run them in reasonable time frames.

This was partially mitigated by upgrading the Django framework to a more recent version (over the lifetime of the project, from 1.6 in 2014, to 1.9 in 2016 and 1.10 in 2017 for final analysis), which allowed the use of the full_text search capability of the underlying PostgreSQL database[32]. Although this eventually improved the retrieval speed (reducing it by around 50%), it did not significantly improve count and aggregation operations (such as counting the number of times that a word appeared in a given volume of tweets), and initial operations after the installation were slower due to the system needing to build to_tsvector and plain to_tsquery objects in the database for millions of rows and objects. As an aside, combining these search methods (regular full table lookup and search vector structures) was slower by an order of magnitude, taking up to 8 minutes to run.

The analysis utilised was a compromise measure, driven by the limits of computational power and optimised code available during the life of the research project. Two key assumptions were made. The first was that the widest set of tweets gathered was a reflective set of the overall data available through whatever sampling method twitter made available through the API, and that these tweets were a representative sample of the overall discourse available.

32

The gathering process introduced a number of sampling biases, which it was difficult to correct after the fact. As a result, although the results shown here – that the protests come to dominate the discourse over the time – are interesting (although perhaps not particularly startling), it is difficult to draw any direct conclusions on the relationships that may exist between policy and the public discourse here. Instead, the analysis here will examine the findings from engaging with Twitter (and social media more generally) as a data source.

Review of Fieldwork Capabilities and Future Direction

> "*However, due to various limitations, qualitative data analysis software should rather be used as a supportive tool than a product that drives the whole research process. In the end, the interpretation of the findings still has to be done by the researcher.*" (Einspänner, Dang-Anh, and Thimm, 2014:98)

This comment from Einspänner, *et. al.* rings painfully true in the case. Although the decision to oversample the data was considered, and deliberate, it may have rested on an improper selection process, and a naive assumption that it would be possible to correct or address any issues with the data set through programmatic analysis. As this was not the case, the rest of this chapter focuses on the specific experience of working with Twitter and social media data, and considers how the approach could be better managed for future research projects.

Twitter as Multi-Level Communication

Bruns and Moe (2014) argue convincingly that Twitter data should be considered in multiple streams. It is clear from the failures of analysis in this research that treating the data as one homogenous source is not a viable option, particularly when that data is gathered together from disparate keyword and geographic searches.

They offer a model of Twitter communication that falls into three layers - micro, meso and macro. Micro communications are direct messages exchanged between users of the site, which are private to them and not visible to other users, and are analogous to emails. The meso level of communication is the follower-followee relation, where users see the feeds of people that they have subscribed to follow, and is the dominant layer of communication on the site. Finally, the macro level is through the use of hashtag based exchanges, where people participate in conversations that are visible site wide to anyone clicking in to a hashtag.

This straightforward delineation of the potential layers of communication on the site would have worked well within the research. The data gathering and analysis did not take into account any significant concept of there being a range of purposes of communication on the site. In particular, a key approach now would be to focus research around hashtags, rather than keywords. This appears to have been a significant mis-targeting in the research design. Bruns and Moe note that "*logically, it is considerably more difficult to move beyond the relatively well-behaved confines of macro-layer hashtag studies.*" (*ibid*:24) Using a hashtag based search approach for data gathering would also be consistent with the many-to-whole aspects of assemblage theory; in applying a hashtag to their content, individuals are opting in to a wider discourse with others that they may not know or follow directly.

In addition to considering this approach to tiering data from Twitter, it would also be useful to consider the long tail model: "*a comparatively small number of highly active users are likely to dominate the dataset… [with] a much larger 'long tail' [of less active users].*" (Anderson, 2006) This same phenomenon was noted in the analysis, but was not capitalised on. As noted in Chapter Seven , with over 7m tweets captured from just over 1.3m unique users, around 2m of those tweets were accounted for by the top 1% most frequent users. This would have reduced the dataset for analysis by 71%, and allowed a concentration of effort on those 13,000 prolific users who were dominating the conversation.

This would also have had the effect of obviating some of the difficulties that arose around the volume and pervasiveness of data noise, and reduced the computational complexity of running pairwise comparisons on tweets as discourse fragments. In future work, it would be possible to run the analysis by utilising these users as the spine of the conversation over time, and utilising retweets, replies and mentions as the responses, for the purposes of comparison. This represents an extremely promising and useful approach for the future.

Threading and Graphing of Networks

One of the main disappointments having gathered this large and promising data set was the unavailability of conversation or 'threading' data from the tweets themselves. As discussed in the analysis chapter, this was a feature that had made its way to the web interface for users, but was not available in the API at the time of researching; even revisiting the individual tweets a year later showed that this information was not available for 'legacy' tweets. However, the concept of being able to show a graph network of the relationship between tweets, with popular tweets as central nodes, and potentially weighting the centrality and/or importance of those nodes against the number of followers that a user has, would be interesting – doubly so if this can be shown over the course of time, perhaps as an animated visualisation of those conversations over a specified time period.

Considering Twitter as Interdependent

> "*To fully understand information flows not just on, but through Twitter as a communicative tool, these outside layers must also be taken into account... Information flows weaved in and out of Twitter, and across the three communicative layers, multiple times. To examine such complex processes of information dissemination only from the perspective of any one layer, or even of any one medium, is to miss an*

> *important dimension of their communicative dynamics*" (Bruns and Moe, 2014:24)

This quote illuminates one of the key limitations of this study. Twitter is one aspect of discursive assemblages, but the fieldwork and analysis neglected to consider any other aspects. By limiting the analysis to solely the textual data retrieved from Twitter, a number of layers of media have been overlooked. These include video content, pictures (which can be easily embedded in tweets), as well as the content of external sites users may link to. Twitter does not exist in splendid isolation for anyone, least of all its users, but is rather part and parcel of individuals' days, as well as of media coverage. Mainstream media and social media have a reciprocal arrangement, with social media commentary on news articles and video clips, as well as tweets being featured in news stories themselves.

As a suggestion for further work, it would be illuminating to use this new technique to examine the relationships and coalescence of an issue utilising the news stories as a kind of anchor. Given the structure of the tweet data outlined in Chapter Seven, links are easily gathered as part of the dataset when tweets are captured through the data streams. This approach would significantly improve the quality and relevance of the data. In this case of this research, this would have elevated the fieldwork from a simple and unsatisfactory analysis of the raw numbers, hamstrung by the mass of its own data, to one that considered a number of textual sources. Being able to draw comparisons of sentiment between groups in relation to a particular article or story would allow evaluation of particular stories and news articles as key nodes in the discussion itself. Although they are not hashtags, URLs are easily identifiable, and can be used as unique identifiers. In terms of the macro space discussed above, the presence of URLs and its accompanying material would contextualise tweets, as well as providing entry points for gathering data about protests, surveillance or any other topic of interest. Such an approach also allows an examination of the centrality of particular stories or sources to tweet networks.

This could be linked back to the approach outlined in the previous section, where prolific users would be taken to be the backbone of any particular conversation. Utilising a combination of both approaches would represent a highly targeted approach to understanding the shape, sentiment and velocity of discourse over time.

Metrics and Objective setting in field work on Social Media

A significant issue in terms of the transition from data gathering to fieldwork is that there were no measurable metrics defined, either in a qualitative or quantitative sense, that would allow the researcher to know that the research was on the right track. In some respects, this led to something of a catchall approach, with the naive assumption that these metrics or insights could be reached inductively, despite the considerable effort expended developing a substantial framework in earlier chapters. Bruns and Stieglitz (2014) offer a number of approaches to considering the metrics used for understanding Twitter. They also note that effective use of such (any) metrics "*also depends on a deeper understanding of the communicative phenomena which they describe... a focus merely on the raw figures themselves is likely to obscure more important patterns within the data.* " (*ibid:72)*

In addition to the basic structural aspects of tweets, as outlined in Chapter 7, they consider the possibility of temporal metrics and group based metrics. Temporal metrics include tweet volume, volume over time, and the breakdown of original tweets, retweets, and replies, or the presence of URLs in tweets. These ratios of original content to retweeted content, as well as URLs and hashtags, can be used to track the dissemination of key information and concepts through the platform. Group metrics have been covered in the section discussing twitter as a site of multi-layered communication.

As these metrics were not identified at any stage in the process - before data gathering, before refocusing to a different form of analysis, or before conducting the analysis - there were no clear criteria for success of the

analysis, which proved a fundamental mis-step. The ideal time to consider the metrics for success of fieldwork and analysis would have been at the point the research questions were reformulated.

Part Four: Conclusions and Further Work

Chapter 9: Reconsidering Theory, Conclusions and Further Work

Introduction

This chapter considers the research questions of the thesis and re-examines the theoretical approaches laid out in Part Two of the thesis in the face of both the difficulties faced in creating an operational approach, and of the insights gained through that research. It argues that the theoretical framework remains a potentially significant tool for generating insight into ongoing social and political issues, and highlights where additional work would allow more specific focus that can provide more comprehensive understanding of those issues, and how this can better feed into policy-making, particularly around surveillance. Crucially, it also offers specific and actionable approaches that - with the benefit of hindsight - could have been taken to mitigate some of the subsequent problems discovered in the fieldwork and analysis stages. It then goes on to advocate for the usefulness of such approaches in future research.

Revisiting Aims and Objectives

- What discursive assemblages are in place around surveillance (particularly at protests), and how do they persist and replicate across different contexts?
- What are the characteristics of those discursive assemblages and how do they change over time in relation to 'real-world' events?
- How can existing social theory be utilised to adequately analyse and provide insight into the operation of assemblages?
- What does operationalising the assemblage in this way have to offer in terms of critical understanding of public policy and surveillance?

Presence of discursive assemblages

It is evident from the sheer volume of data generated and captured through the fieldwork methodology that there are large-scale discourses on social media created and sustained around protests. Previous sections have noted the significant difficulty in finding and isolating discourses specifically of surveillance, as had been the original hypothesis.

Operationalising the Assemblage - Recommendations

As was shown in Chapter Eight, although the research was unsuccessful in this particular case, there are an important number of recommendations that can be made to future scholars in terms of operationalising this type of approach. Some are generally good research practice, that it is tempting to abandon in the face of this bold new theoretical approach, while some are more specific.

The first is that, as has been emphasised and argued repeatedly throughout the research, but ultimately not heeded in the fieldwork carried out, assemblages are complex, volatile structures. By definition, they are hard to describe in existing terms, and rightly defy an essentialist description. With this in mind, research projects attempting to capture and understand the relationships within an assemblage should start with a small example. Once this proof of concept has been successfully completed, the research can be scaled, but it is all too easy for research using programmatic methods to quickly collapse under the weight of their own data.

This notion of success quickly leads to the second recommendation: metrics for success should be clearly defined. If they are unreachable for analytical or data quality reasons, they should be quickly reconsidered and the data gathering or analysis retested. Although it may not be necessarily clear how the data will answer the specific research question at the outset, there should be a rough and indicative path as to how to move from the hypothesis,

through the data and towards a stage where that will become clearer. Fundamentally, this research failed to move past this point, due to the lack of success criteria between the data and the original research questions.

Related to both of these recommendations is that there are a number of options to make the sample size smaller and more appropriate, particularly in the early stages of testing an approach. Although all research is iterative, the feedback times using programmatic research for social sciences can be shorter than may be the case in more traditional fieldwork methods; the process can be more akin to a live literature review than to conducting interviews or media framing exercises.

The tendency to consider social media as entirely reflective of an issue should generally be rejected, unless it is the sole object of research. Social Media is one aspect of discursive emanations, and care should be taken to consider where and how it interacts with other media; at the same time as the media narrative shapes discussions on social media, social media has increasingly been shaping media narratives, through reaction and virality.

In terms of programmatic research, the value of writing tests cannot be overestimated. Many of the bandwidth issues with the program in the early stages of the research could have been avoided by writing unit and integration tests for the system, to mimic the high traffic and high throughput scenarios that it saw in practice. Instead, there was a period of live trial and error, which caused gaps in data coverage, as well as a degree of manual rework and toil to check system uptime, availability and coverage.

Policy as an Assemblage

At the outset, the thesis was that:

> "*Rather than being reliant on domain-specific heuristics, as has traditionally been the case, assemblage theory in public policy provides*

a context-sensitive, scale appropriate system of analysis that can operate across domains and jurisdictions."

Analysis to generate insight into this position was severely hindered by the performance issues of the software. The assumption that overcapture would be a viable strategy in terms of gathering data was valid; it was not valid for the purposes of analysis. This has consequently hindered the connection of the fieldwork to the theoretical framework laid out in Chapter Five.

From the limited analysis that it was possible to carry out, there is a strong indication that the broad approach was correct: discursive assemblages territorialised around a particular topic, and the next stage would be to understand the sentiment around those topics. This would allow the discursive assemblants to be understood as vectors, with strength and direction. In turn, widening the comparators for discursive comparisons and programmatic LSA may allow us to bridge the gap between micro-discourse as expressed through social media, and the macro-discourse – the process of tracing a path through those discourses, but only if the warnings and findings outlined in previous chapters are heeded.

Policy documents and news stories can both be considered as nodes in the discursive network, which are then shared, discussed and argued about on social media. These sharing and discussing actions, carried out by the general public as well as by journalists and political commentators, provide that elusive set of connections that is missing from having only analysed the disparate and spread discursive assemblages of the public.

"*The complex relations between policy formulation and implementation may be interpreted as implying that the two activities form a seamless web: an interpretation that would bring us back to a monolithic concept. But can also be modeled as a nonlinear process involving feedback, a process of formulation-implementation-reformulation that does not jeopardize the 'ability to assess the extent of goal attainment and the*

distribution of authority between elected and appointed officials'."
(DeLanda, 2006:85)

Considering policy as a complex and emergent system is more reflective of the observable chaos in the modern political world. Taking the stance that governments and policy do less governing than is generally assumed seems an increasingly valid world view. The consequential question is how to use that to create better policy? As ever, the question of 'better' policy is strongly contested, but an understanding of the constitutive assemblages that surround policy networks may yet facilitate a 'better understanding of "better"' amongst the policy community. The remaining contributor of complexity is, and will remain, the political elements, which constitutes an unhelpful line of flight and competing (and sometimes counter) rationality to the policy world.

As to the concept of 'better policy' – only broad recommendations can be made at this stage, for example that policy cycles (and other heuristics) should be shorter, and more iterative. Dramatic policy interventions should be thought of more as a line of flight: understanding that deterritorialising existing institutions and assemblages may be necessary to effect significant change. The practical mechanisms to achieve this are contextual, but rethinking 'success' will be a necessary precursor to that.

The Actual and the Virtual

During the course of the research, the parallels between Haggerty's data doubles – in effect, memory and retention of the assemblage – and Deleuze's consideration of how the actual and virtual coexist emerged repeatedly. The surveillant assemblage is an iconic concept in Surveillance Studies, and one that has solidity in terms of both the data that represents it and in its material effects on the world and the social sorting of its real life parallel. However, it also has interesting ontological concepts when considered in the light of the more in depth theory of assemblages:

> *"Hence there is a coalescence and division, or rather oscillation, a perpetual exchange between the actual object and its virtual image: the virtual image never stops becoming actual… This perpetual exchange between the virtual and the actual is what defines a crystal; and it is on the plane of immanence that crystals appear. The actual and the virtual coexist, and enter into a tight circuit from which we are continually retracing from one to the other. This is no longer a singularization, but an individuation as process, the actual and its virtual: no longer an actualization but a crystallisation."* (Deleuze and Parnet, 1987)

It is accepted that data doubles are actual. They are often 'more real' than the people they represent, but this leads to a question about the nature of truth in surveillance. The crystallisation of an assemblage, whether around the actions to be taken on an individual as a result of a surveillant sorting, or of a policy assemblage at a higher level, may lead to 'facts' or truth that is objectively wrong. The subjective truth brought to bear on the surveilled body diverges from the objective truth that is derivable standing apart and outside the assemblage. Unfortunately for the surveilled individual, this is difficult, if not impossible. However, it is an important ontological question – Deleuze argued that the virtual no longer needed to actualize itself. It is actual in itself, and is *in any case* such a close corollary to the actual that the two are indistinguishable. If surveillance has become one of the dominant modes of organisation for modern society, how then to treat this ontological dichotomy?

We have created a system (state, policy and surveillant alike) that allows the real and the virtual to depart. Most people or surveillant subjects are uninteresting to the systems and functions that can process data, but the tangential information we can attach to their likeness is not - a significant volume of the data generate by these systems is generated by machines for the sake of other machines. As we increasingly rely on both praxis of surveillance and machine based data generation for our organisational and

social construction, the question of how we reconcile the increasing distance between the real and the virtual is a question that bears significant thought.

Contributions to Knowledge - Research Methodology

This research has generated and tested an novel research methodology in showing that large-scale research projects around social media are mostly practical on a single researcher basis. The concept of 'doing surveillance studies' by effectively carrying out surveillance is also worthy of further discussion and exploration; there is no small degree of irony in having carried out this project by becoming a node of surveillance. On the face of it, this is justified by both the attempts to understand the surveillant and policy assemblages in place in the UK, and also in trying to understand how those can best be destabilised and disrupted. Perhaps, as a community, surveillance scholars may only be able to, if they desire, disrupt the power flows of surveillance by adopting, or co-opting some of them for ourselves.

As a methodology, the analysis of Twitter was well intentioned, but flawed. Although it allowed a large volume of data to be collected, it was poorly targeted, with consequent problems in analysing and managing this data. It was also, despite its broad reach, too shallow. Tweets are such fragmentary devices that without larger 'anchor' nodes to compare against, for example, newspaper articles or public speeches, the value and insight that can be gleaned is questionable at present. However, as a newer type of analysis, it bears further work and examination.

Reflections and Directions for Future Research

> "The limitations of social scientific research based on Twitter data stem from constraints which impact research projects on different levels. As with any other methodology, not all types of data and forms of analysis align themselves equally well with all kinds of research questions." (Gaffney and Puschmann, 2014:64)

Gaffney and Puschmann capture the central issue with the research well here with this statement. There was a significant effort to use Twitter data, in isolation, to answer difficult questions of social structures that this research acknowledges are amongst the most complex ways to conceptualise society. Specifically, it attempted to answer qualitative questions about how people felt and expressed themselves around protests and surveillance by using quantitative analyses. Any future research must lean more heavily on the guidance of the researcher, and ask questions more closely aligned to the topic at hand than the generalised ones asked here. Specifically, that guidance and reliance on the work of the researcher must focus on the *interpretation* of the data generated, through reference to previous work, even where a study departs from the existing literature. Gaffney and Puschmann note that:

> "*...An ideal study should be well grounded in a specific set of research questions and query the data in accordance with them. In contrast to traditional instruments such as surveys and conventional content analysis, it is important to note that even the exploratory phase of research is markedly quantitative when exploring social media. Since searching, filtering, and ranking are the only feasible way to make masses of content readable to the human researcher, they form a logical first step in any analysis, even in qualitative studies*. " (*ibid,* 64)

Although they argue for the 'markedly quantitative' nature of the research here, a key finding from this research is that, in interpreting assemblages and discourse, it is extremely difficult, and often undesirable to separate the quantitative from the qualitative. Using this type of data, particularly at any kind of larger scale, the quantitative *becomes* qualitative. The volume of discourse on a topic carries its own implications, as do the tacit endorsements or criticisms that come from retweeting or replying to someone else's content.

Societal reflections

The central tenet of this thesis research was that it would be possible to take a small corner of the discursive sphere, around a specific event or set of events, and map the discourse and interactions happening around them.

What has been uncovered through the course of this research is that as researchers, policymakers or even as activists, we are ill-equipped to handle the pace, scale and complexity of the new media world enabled through social media. As a conceptual framework, the concept of assemblages of government, public policy and subjectivity are compelling. As an implementable research methodology, however, it has some further work required before more generalisable concepts and methods can be taken to be widely applicable and of utility to researchers.

The first entry into that opening should necessarily be understanding how discursive assemblages can be quantitatively assessed in something approaching a real-time basis. The primary challenges from that are firstly around ethical issues, secondly about the trade offs between simple brute force computing complexity and good research practice, and lastly about the what notion of public good it can serve.

For policymakers, the potential of theorising policy as an assemblage is significant, but comes with a significant overhead: as some of the primary beneficiaries from the metastability of government, they also have potentially the most to lose by changing how they theorise about their own relationships with power and the structures that maintain it. The interface of public discourse to policy is strongly defended by its gatekeepers, meaning that in practice, the policy process may simply choose to disregard those inputs that it finds distasteful, or that would deterritorialise assemblages that suit the existing holders of power.

In terms of critical utility, assemblage theory has a similar potential for activists and protestors; the potential to disrupt existing structures and

address asymmetries of power is significantly enhanced by pursuing these approaches. This comes with an important caveat that the field of policy assemblage studies needs to be significantly more advanced before being used as the basis for action, lest existing assemblages be broken apart and reform in new, and more onerous configurations.

Bibliography

Ackoff, R.L. (1974) *Redesigning the Future: Systems Approach to Societal Problems.* 1st ed. John Wiley & Son.

Albrechtslund, A. (2007). Ethics and technology design. *Ethics and information technology*, *9*(1), 63-72.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. (2001) Fractionalization. *Journal of Economic Growth.,* 8 (2), pp. 155-194.

Allison, G. T. (1971). *Essence of decision: Explaining the Cuban missile crisis* (No. 327.5 (729.1)). Little, Brown and Company.

Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. Hachette Books.

AoIR, Markham, A. and Buchanan, E. (2012) *Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee.* 2. Online: Association of Internet Researchers. Available: http://aoir.org/reports/ethics2.pdf [Accessed: 2016-02-02].

Bacchi, C. (2000). Policy as discourse: What does it mean? Where does it get us?. *Discourse: studies in the cultural politics of education*, *21*(1), 45-57.

Bachrach, P. and Baratz, M. (1970) *Power and poverty: Theory and practice.* Oxford: Oxford University Press.

Ball, K. and Haggerty, K.D. (2005) Doing Surveillance Studies: Editorial. *Surveillance and Society,* 3 (2.3), pp. 129-138.

Ball, K. (2009). Exposure: Exploring the subject of surveillance. *Information, Communication & Society*, *12*(5), 639-657.

Barnard-Wills, D. (2011) UK News Media Discourses of Surveillance. *The Sociological Quarterly,* (4), pp. 548.

Barnard-Wills, D. (2012) *Surveillance and identity: Discourse, subjectivity and the state*. Farnham and Bulington VT: Ashgate.

Barnett, C., Cloke, P., Clarke, N. and Malpass, A. (2011) *Globalizing Responsibility: The Political Rationalities of Ethical Consumption.* Chichester: Wiley Blackwell.

Bauman, Z., Bigo, D., Esteves, P., Guild, E., Jabri, V., Lyon, D., & Walker, R. B. (2014). After Snowden: Rethinking the impact of surveillance. *International political sociology, 8(2), 121-144.*

BBC, 2011, *Anti-cuts march: Tens of thousands at London protest,* Available: https://www.bbc.co.uk/news/uk-12864353 [Accessed: 2019-06-06]

BBC, 2012, *Anti-austerity marches take place,* Available: https://www.bbc.co.uk/news/uk-20007496 [Accessed: 2019-06-06]

Becker, G.S. (1968) Crime and Punishment: An Economic Approach. *Journal of Political Economy,* 76 (2), pp. 169-217.

Becker, G. S., Murphy, K. M., & Tamura, R. (1990). Human capital, fertility, and economic growth. *Journal of political economy*, *98*(5, Part 2), S12-S37.

Bennett, A., & George, A. L. (2005). Case studies and theory development in the social sciences. *Cambridge, Mass*.

Bergson, H., (1910) "The Multiplicity of Conscious States; The Idea of Duration". Chapter 2 in *Time and Free Will: An Essay on the Immediate Data of Consciousness*, translated by F.L. Pogson, M.A. London: George Allen and Unwin pp. 75-139.

Birkland, T.A. (2011) *An introduction to the policy process : theories, concepts, and models of public policy making.* New Yorks: Armonk.

BMJ 2001;323:1103: Eysenbach, G., & Till, J. E. (2001). Ethical issues in qualitative research on internet communities. *Bmj*, *323*(7321), 1103-1105.

Bogard, W. (2006). Surveillance assemblages and lines of flight. In *Theorizing surveillance* (pp. 111-136). Willan.

Bonta, M., & Protevi, J. (2004). *Deleuze and Geophilosophy: a guide and glossary.*

Bourdieu, P. (1998) *Practical Reason.* Stanford: Stanford University Press.

Bové, P., Lentricchia, F., & McLaughlin, T. (1990). Critical terms for literary study. *Discourse*, 50-65.

Bruns, A., & Stieglitz, S. (2014). Metrics for understanding communication on Twitter. *Twitter and society*, *89*, 69-82.

Bruns, A., & Moe, H. (2014). Structural layers of communication on Twitter. In *Twitter and society* (Vol. 89, pp. 15-28). Peter Lang.

Bundgaard, U. and Vrangbaek, K. (2007) Reform by Coincidence? Explaining the Policy Process of Structural Reform in Denmark. *Scandinavian Political Studies,* 30 (4), pp. 491-520.

Cabinet Office (2006) *Executive Agencies: A Guide for Departments.* London: HMSO.

Campbell, C. and Van Brakel, R. (2015) Privacy as a Line of Flight In Societies of Mass Surveillance. *Ethical Space: International Journal of Communication Ethics,* 12 (3/4), pp. 39-46.

Castells, B., 2012, *Networks of Outrage and Hope: Social Movements in the Internet Age*, Cambridge: Polity

Cesare, N., Grant, C. and Nsoesie, E.O. (2017) *Detection of User Demographics on Social Media: A Review of Methods and Recommendations for Best Practices.* Cornell University: Cornell University. Available: https://arxiv.org/abs/1702.01807v1 [Accessed: 2017-06-01].

Chu, Z., Widjaja, I. and Wang, H. (2012) Detecting Social Spam Campaigns on Twitter. In: (eds), vol 7341. Springer, Berlin, Heidelberg. In: F. Bao, P. Samarati and J. Zhou, ed. *Applied Cryptography and Network Security..ACNS 2012.* HeidelbergBerlin: Springer.

Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM, 31(5), 498-512*.

Cohen, BC. (1963). *The press and foreign policy*. Princeton, NJ: Princeton

Collier, D. (2011) Understanding Process Tracing. *PS: Political Science and Politics, 44 (4), pp. 823-830.*

Considine, M. (2005) *Making public policy : institutions, actors, strategies.* Cambridge: Polity.

Daguerre, A. (2004). Importing workfare: policy transfer of social and labour market policies from the USA to Britain under New Labour. *Social Policy & Administration*, *38*(1), 41-56.

Dahl, R. A. (1961). *Who governs? democracy and power in an american city* (No. 320.973 D3).

Dandeker, C. (2006) Surveillance and Military Transformation: Organizational Trends in Twenty First Century Armed Services. In: Haggerty, K. and Ericson,R., eds. *The New Politics of Surveillance and Visibility.* Toronto: University of Toronto Press, pp. 225-249.

Dandeker, C. (1990) *Surveillance, power and modernity : bureaucracy and discipline from 1700 to the present day.* Cambridge: Polity.

David, B., & Lindblom, C. E. (1963). *A Strategy of Decision: Policy Evaluation as a Social Process.*

Davis, M. (1990) *City of Quartz*, London: Pimlico.

Davis, M. (1992) 'Fortress Los Angeles: The Militarization of Urban Space', in Sorkin (ed.) *Variation on a Theme Park*, New York: Hill and Wang.

Dean, M. (1999) *Governmentality : power and rule in modern society.* London: SAGE.

Dean, M. (2010) *Governmentality : power and rule in modern society.* 3rd ed. London: SAGE.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R. (1991) Indexing by latent semantic analysis. *Journal of the American Society for Information Science,* 41, pp. 391-407.

DeLanda, M. (1997). *A thousand years of nonlinear history,* London: Zone Books

DeLanda, M. (2006) *A New Philosophy of Society: Assemblage Theory and Social Complexity.* 1st ed. London: Continuum.

Deleuze, G. (1968), *Difference and Repetition*, 2nd Edition (2014), London: Bloomsbury

Deleuze, G. and Guattari, F. (1987) *A thousand plateaus: capitalism and schizophrenia.* London: Continuum.

Deleuze, G., (1988), *Foucault*, Minneapolis: University of Minnesota Press

Deleuze, G. (1991) *Empiricism and Subjectivity.* New York: Columbia University Press.

Deleuze, G. (1992) Postscript on the Societies of Control, *October*, 59, 3-7. Retrieved July 7, 2019, from www.jstor.org/stable/778828

Deleuze, G. (2007). *Two regimes of madness: Texts and interviews 1975-1995*, Boston: MIT Press

Deleuze G and Parnet C (1987) *Dialogues II*, New York NY, Columbia University Press

Dencik, L., Hintz, A., Carey, Z., & Pandya, H. (2015). *Managing 'threats': uses of social media for policing domestic extremism and disorder in the UK.*

Derrida, J. (1981). Interview with Julia Kristeva. *Positions, 15-36.*

Dosse, F. (2012) Deleuze and Structuralism. In: D.W. Smith and H. Somers-Hall, eds. *The Cambridge Companion to Deleuze.* 1st ed. Cambridge: Cambridge University Press, pp. 126-150.

Douglas, M. (1986). *How institutions think*. Syracuse University Press.

Dror, Y. (1986) *Policymaking under adversity.* New Brunswick: Transaction Publications.

Dryzek, J. S. (1993). Policy Analysis and Planning: From Science to Argument. The Argumentative Turn in Policy Analysis and Planning, 213.

Dumais, S.T., Littman, M.L. and Landauer, T.K. (1998) Automatic cross-linguistic information retrieval using latent semantic analysis. In: G. Grefenstette, ed. *Cross-language information retrieval.* Boston: Kluwer, pp. 51-62.

Dumais, S.T. (2004) Latent semantic analysis. *Annual Review of Information Science and Technology,* 38 (1), pp. 188-230.

Easterby-Smith, M. T., & Thorpe, R. (1991). R. and Lowe, A.(1991). Management research: An introduction, 1.

Einspänner, J., Dang-Anh, M., & Thimm, C., (2014). Computer-assisted content analysis of Twitter data. *Weller K., Bruns A. Burgess J., Mahrt M.(eds.), Twitter and Society, New York, Peter Lang*, 97-108.

Enserink, B., Koppenjam, J.F.M. and Mayer, I.S. (2013) A Policy Sciences View on Policy Analysis. In: W.A.H. Thissen and W.E. Walker, eds. *Public Policy Analysis : New Developments.* New York: Springer US, pp. 11.

Entman, R. M. (2003). Cascading activation: Contesting the White House's frame after 9/11. *Political Communication,*, *20*(4), 415-432.

Ericson, R., and Haggerty, K. D. (Eds.) (2006). *The new politics of surveillance and visibility.* Toronto: University of Toronto Press.

Etzioni, A. (1967) Mixed-Scanning: A "Third" Approach to Decision-Making. *Public Administration Review,* 27 (5), pp. 385-392.

Exworthy, M. and Powell, M. (2004) Big Windows and Little Windows: Implementation in the Congested State. *Public Administration,* 82 (2), pp. 263-281.

Eysenbach, G. and Till, J.E. (2001) Ethical issues in qualitative research on internet communities. *Bmj,* 323, pp. 1103.

Faden, R.R. and Beauchamp, T.L. (1986) *A HIstory and Theory of Informed Consent.* Oxford: Oxford University Press.

Ferlie, E., Pettigrew, A. and Ashburner, F.L. (1996) *The New Public Management in Action.* Oxford, UK: Oxford University Press.

Fischer, F. and Forester, J. (1993) *The Argumentative Turn in Policy Analysis and Planning.* Raleigh, NC.: Duke University Press.

Flanagin, A., & Metzger, M. J. (2017). Digital media and perceptions of source credibility in political communication. *The Oxford handbook of political communication*, *417*.

Foley, D. K. (2003). Endogenous technical change with externalities in a classical growth model. Journal of Economic Behavior & Organization, 52(2), 167-189.

Fornet-Betancourt, R., Becker, H., Gomez-Müller, A., and Gauthier, J.D. (1987) The Ethic of Care for the Self as a Practice of Freedom: An Interview with Michel Foucault on January 20, 1984. *Philosophy & Social Criticism, 12(2–3), 112–131.*

Foucault, M. (1991) Governmentality. In: G. Burchill et al., eds. *The Foucault Effect.* Chicago: University of Chicago Press.

Foucault, M. (1997) Security, Territory and Population. In: P. Rabinow, ed. *Michel Foucault, Ethics: Subjectivity and Truth.* New York: The New Press, pp. 67-71.

Foucault, M. (2007) *Security, territory, population : lectures at the College de France, 1977-78.* Basingstoke: Palgrave MacMillan.

Foucault, M., Davidson, A. I., & Burchell, G. (2008). *The birth of biopolitics: lectures at the Collège de France, 1978-1979*. Springer.

Frois, C. (2011) Video SurVeillance in Portugal: Political Rhetoric at the Center of a Technological Project. *Social Analysis,* 55 (3), pp. 35-53.

Fuchs, C. (2011) *Foundations of Critical Media and Information Studies.* [Routledge Advances in Sociology No. 52.] 1st ed. New York: Routledge.

Füredi, F. (2009) Precautionary culture and the rise of possibilistic risk assessment. *Erasmus Law Review,* 2 (2), pp. 198.

Gaffney, D., & Puschmann, C. (2014). Data collection on Twitter. *Twitter and society*, 55-67.

Galtung, J., (1969), Violence, Peace, and Peace Research, *Journal of Peace Research*, 6(3), pp. 167-191, https://doi.org/10.1177/002234336900600301

Gandy, O.H. (1993) *The Panoptic Sort: A Political Economy of Personal Information.* Boulder, CO: Westview Press.

Gastil, J. (2008). *Political communication and deliberation*. Sage.

Guardian (2011), *UK riots: Mobile phone operators help police to identify looters,* https://www.theguardian.com/uk/2011/aug/09/uk-riots-mobile-phone-operators [Accessed: 2019-04-16]

Guardian (2014), *Facebook receives nearly 2,000 data requests from UK police,* https://www.theguardian.com/technology/2014/apr/11/facebook-2000-data-requests-police [Accessed: 2019-05-05]

Giddens, A. (1985) *A Contemporary Critique of Historical Materialism: Vol. 2. The Nation State and Violence.* Cambridge: Polity Press.

Giddens, A. (1990) *The Consequences of Modernity.* New Ed edition (April 1991) ed. Polity Press.

Giddens, A. (1991) *Modernity and Self-Identity: Self and Society in the Late Modern Age.* Stanford University Press.

Giddens, A. (1994) *Beyond Left and Right: The Future of Radical Politics. .* 1998 edition ed.

Giddens, A. (1998). The Third Way, *Polity*, *129*.

Gill, J., & Johnson, P. (2002). *Research methods for managers*. Sage.

Gordon, I., Lewis, J. and Young, K. (1977) Perspectives on Policy Analysis. In: S.P. Osborne , ed. *Public Management: Policy making, ethics and accountability in public management.* Vol V ed. Routledge, pp. 11.

Green, D. and Shapiro, I. (1996) *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science.* Boston: Yale University Press.

Grimmelman, J. (2015) The Law and Ethics of Experiments on Social Media Users. *Colorado Technology Law Journal,* 219 (13).

Gross, J. L., & Yellen, J. (2005). *Graph theory and its applications*. CRC press.

Guardian, (2010) *Student protests - as they happened.* Available: https://www.theguardian.com/education/blog/2010/dec/09/student-protests-live-coverage?INTCMP=SRCH [Accessed: 2019-06-20]

Habermas, J., (1989), *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*, Cambridge Massachusetts: The MIT Press

Haggerty, K. D., & Ericson, R. V. (1999). The militarization of policing in the information age. *Journal of Political and Military Sociology, 27(2), 233*.

Haggerty, K.D. and Ericson, R.V. (2000), The surveillant assemblage. *The British Journal of Sociology*, 51, pp. 605-622.

Hainmueller, J., Hiscox, M.J., Hainmueller, J. and H. (2010) Attitudes toward Highly Skilled and Low-skilled Immigration: Evidence from a Survey Experiment. *The American Political Science Review.,* 104 (1), pp. 61-84.

Hall, P.A. (1992) The movement from Keynesianism to monetarism: Institutional analysis and British economic policy in the 1970s. *Structuring Politics: Historical Institutionalism in Comparative Analysis,* , pp. 90-113.

Ham, C. and Hill, M. (1984) *The Policy Process in the Modern Capitalist State.* 1993 ed. Hemel Hempstead: Harvester Wheatsheaf.

Hand, M. (2014) From cyberspace to the dataverse: Trajectories in digital social research. *Big Data? Qualitative Approaches to Digital Research.* Emerald Group Publishing Limited, pp. 1-27.

Hankey, B.F., Ries, L.A., Edwards, B.K. (1999) The Surveillance, Epidemiology, and End Results Program, Cancer Epidemiol Biomarkers Prev Dec 1 1999 (8) (12) pp. 1117-1121

Heiskanen, B. (2017) Meme-ing Electoral Participation. *European Journal of American Studies,* 2 (12).

Hier, S. and Greenberg, J. eds. (2007) *The Surveillance Studies Reader.* London: Open University Press.

Hill, M. (2009) *The Public Policy Process.* 5th ed ed. London: Longman.

Hogwood, B.W. and Gunn, L.A. (1984) *Policy analysis for the real world.* Oxford: Oxford University Press.

Holland, E.W. (1991) Deterritorializing "Deterritorialization": From the "Anti-Oedipus" to "A Thousand Plateaus". *Substance,* 20 (3), pp. 55-65.

Hood, C. (1990) De-Sir Humphreyfying the Westminster Model of Bureaucracy: A New Style of Governance? *Governance,* 3 (2), pp. 205-214.

House of Commons, (1854), Report on the Organisation of the Permanent Civil Service, q/JN 426 NOR Back

Howlett, M. (2009) *Studying public policy : policy cycles & policy subsystems .* Oxford: Oxford University Press.

Immergut, E. M. (1992). *Health politics: interests and institutions in Western Europe*. CUP Archive.

Iversen, T., Soskice, D., Iversen, T. and S. (2006) Electoral Institutions and the Politics of Coalitions: Why Some Democracies Redistribute More Than Others. *The American Political Science Review.,* 100 (2), pp. 165-182.

Iyengar, S. and Hahn, K.S. (2009) Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication.,* 59 (1), pp. 19-39.

Jamieson, K.H. and Cappella, J.N. (2008) *Echo chamber: Rush Limbaugh and the conservative media establishment.* Oxford University Press, USA.

Jenkins-Smith, H. and Sabatier, P. (1993) “The Study of Public Policy Processes.” In P. Sabatier and H. Jenkins-Smith (eds), . Boulder, CO: Westview Press, 1–9. In: P. Sabatier and H. Jenkins-Smith, eds. *Policy Change and Learning: An Advocacy Coalition Approach.* [ails to provide a good vehicle for integrating the roles of policy analysis and policy oriented learning throughout the public policy process.] Westview Press, pp. 1-9.

Jessop, B. (1982) The Capitalist State: Marxist theories and methods, Oxford: Blackwell.

Jessop, B. (2005) *From Thatcherism to New Labour: Neo-Liberalism, Workfarism, and Labour Market Regulation.* http://www.lancs.ac.uk/sociology/research/publications/papers/jessop-from-thatcherism-to-new-labour.pdf.

John W. Kingdon and Kingdon, J.W. *Agendas, alternatives, and public policies / John W. Kingdon.* New York: New York : HarperCollins College Publishers, c1995.

John, P. (2012) *Analyzing public policy .* Abingdon: Routledge.

Jones, C., Hesterly, W.S. and Borgatti, S.P. (1997) A General Theory of Network Governance: Exchange Conditions and Social Mechanisms. *Academy of Management Review,* 22 (4), pp. 911.

Kahneman, D. and Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk. *Econometrica,* 47 (2), pp. 263-291.

Karl, T.L. (1990) Dilemmas of democratization in Latin America. *Comparative Politics,* 23 (1), pp. 1-21.

Kenner, A.M., 2008,. Securing the Elderly Body: Dementia, Surveillance, and the Politics of “Aging in Place”, *Surveillance and Society*, 5(3), pp 252-269

KernalMachine project, (2015), Available at https://github.com/kernelmachine/pyLSA [Accessed: 2016-02-03]

Kingdon, J.W. (1995) 'The Policy Window, and Joining the Streams', in Agendas, Alternatives, and Public Policies, Chapter 8, *International Library of Comparative Public Policy.,* 7, pp. 713-743.

Kleinsman, J. and Buckley, S. (2015) Facebook Study: A Little Bit Unethical But Worth It? *Journal of Bioethical Inquiry,* 12 (2), pp. 179-182.

Kramer, A., Guillory, J. and Hancock, J. (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences,* 111 (24), pp. 8788-8790.

Kress, G. and Koehler, G. (2005) Policy drift: An evaluation of the California Business Enterprise Program . *Policy Studies Journal,* 8 (7), pp. 1101-1108.

Kritzman L (1988) Power and sex: An interview with Michel Foucault. In: Kritzman LD (ed.) *Michel Foucault: Politics, Philosophy, Culture: Interviews and Other Writings, 1977–1984.* New York: Routledge

Kullenberg, C. and Palmås, K. (2009) *Contagionology.* Eurozine. Available: http://www.eurozine.com/contagiontology/ [Accessed: 2015-06-25]

Landauer, T.K., Foltz, P.W. and Laham, D. (1998) An introduction to latent semantic analysis, *Discourse Processes*, 25:2-3, 259-284, DOI: 10.1080/01638539809545028

Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review,* 104(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Lapan, S.D. and Armfield, S.W.J. (2009) Case Study Research. In: S.D. Lapan et al., eds. *Qualitative Research: An Introduction to Methods and Designs.* New Jersey: Wiley.

Lapid, Y., & Kratochwil, F. V. (Eds.). (1996). *The return of culture and identity in IR theory*. Rienner.

Larner, W. (2000) Neo-liberalism Policy, Ideology, Governmentality. *Studies in Political Economy,* 63 (1), pp. 5-25.

Lasswell, H.D. (1951) *The Political Writings of Harold Lasswell.* Glencoe, IL: The Free Press.

Lasswell, H. D. (1955). Current Studies of the Decision Process: Automation Versus Creativity. Western Political Quarterly, 8(3), 381–399. https://doi.org/10.1177/106591295500800305

Latour, B., (2007), Reassembling the Social: An Introduction to Actor-Network Theory, Oxford: Oxford University Press

Lawrence, R.G. (2000) Game-Framing the Issues: Tracking the Strategy Frame in Public Policy News. *Political Communication,* 17 (2), pp. 93-114.

Lemke, T. (2007) An indigestible meal? Foucault, governmentality and state theory. *Distinktion: Journal of Social Theory,* 8 (2), pp. 43-64.

Lemke, T. (2011) *Foucault, governmentality, and critique.* Slough: Paradigm.

Lévinas, E. (1987) *Time and the other : and additional essays].* Pittsburgh, PA: Duquesne University Press.

Lindblom, C.E. (1959) The Science of "Muddling" Through. Emergence : Complexity and Organization., 12 (1), pp. 70-80.

Lomell, H.M. (2004) Targeting the Unwanted: Video Surveillance and Categorical Exclusion in Oslo, Norway. Surveillance and Society, 2 (2/3), pp. 346-360.

Longley, P. A., Adnan, M., & Lansley, G. (2015). The geotemporal demographics of Twitter usage. Environment and Planning A, 47(2), 465-484.

Loveridge, D.J. (1979) Decisions, judgement and style. Long Range Planning, 12 (1), pp. 22-27.

Lubbers, E., (2015) Undercover Research - Corporate and Police Spying on Activists. An Introduction to Activist Intelligence as a New Field of Study. *Surveillance & Society, 13(3/4): pp. 338-353.*

Lupu, N. and Pontsson, J. (2011) The Structure of Poverty and the Politics of Redistribution. *American Political Science Review,* 105 (2), pp. 316.

Lyon, D. (1994) *The Electronic Eye: The Rise of Surveillance Society - Computers and Social Control in Context.* Cambridge: Polity.

Lyon, D. (2001) Facing the future: Seeking ethics for everyday surveillance. *Ethics and Information Technology,* 3 (3), pp. 171-180.

Lyon, D. (Ed.). (2006). *Theorizing surveillance*, London: Routledge.

Lyon, D. (2007) *Surveillance Studies: An Overview.* Cambridge: Polity.

Lyon, D. (2010) Liquid Surveillance: The Contribution of Zygmunt Bauman to Surveillance Studies1. *International Political Sociology,* 4 (4), pp. 325-338.

Lyon, D. (2010) National IDs in a Global World: Surveillance, Security, and Citizenship. *Case Western Reserve Journal of International Law,* 42, pp. 607.

Maass, A., (1966),Benefit-Cost Analysis: Its relevance to public investment decisions, *Quarterly Journal of Economics,* 80(2), May, 208-26

Maass, A., Link, A.N. and Scott, J.T. (1967) `Benefit-Cost Analysis: Its Relevance to Public Investment Decisions', Quarterly Journal of Economics. *The International Library of Critical Writings in Economics.,* (258), pp. 18-36.

Mahrt, M. and Scharkow, S. (2013) The Value of Big Data in Digital Media Research. *Journal of Broadcasting and Electronic Media.,* 57 (1), pp. 20-33.

Majone, G. (1989) *Evidence, argument and persuasion in the policy process.* London: New Haven.

Malins, P. (2004) Machinic Assemblages: Deleuze, Guattari and Ethico-Aesthetics of Drug Use. *Janus Head,* 7 (1), pp. 84-104.

Martin, L.W. and Vanberg, G. (2004), Policing the Bargain: Coalition Government and Parliamentary Scrutiny. *American Journal of Political Science*, 48: 13-27. doi:10.1111/j.0092-5853.2004.00053.x

Marx, G. (1988) *Undercover: Police Surveillance in America.* California: University of California Press.

Marx, G. (2005) Surveillance and Society. *Encyclopedia of Social Theory.*

McBeth, M. K., Jones, M. D., & Shanahan, E. A. (2014). The narrative policy framework. *Theories of the policy process* (3) pp. 225-266.

McConnell, A. (2010) *Understanding Policy Success. Rethinking public policy*, Basingstoke: Palgrave Macmillan

McKee, K. (2009). Post-Foucauldian governmentality: What does it offer critical social policy analysis?. *Critical social policy*, *29*(3), 465-486.

McLaughlin, M.W. (1987) Learning from Experience: Lessons from Policy Implementation. *Educational Evaluation and Policy Analysis,* 9 (2), pp. 171-178.

Michalos, A.C. (1973) Rationality between the maximizers and the satisficers. *Policy Sciences,* 4 (2), pp. 229-244.

Mitraney, D. (1966) *A working peace system.* Pittsburgh, PA: Quadrangle.

Garcia-Molina, H., Ullman, J. D., & Widom, J. (2009). *Database Systems: The Complete Book*.

Moore, M.H. (1995) *Creating Public Value Strategic Management in Government.* Boston: Harvard University Press.

Moore, T., Lapan, S.D. and Quartaroli, M. (2005) Case study research. In: S.D. Lapan and M.T. Quartaroli Riemer, eds. *Qualitative Research.* London: Wiley.

Morgan-Lopez, A.A., Kim, A.E., Chew, R.F. and Ruddle, P. (2017) Predicting age groups of Twitter users based on language and metadata features. . *Plos One 12(8),* 12 (8), pp. 1-12.

Morris, M.H. (2015) *If it's not Scottish, it's Crap!: What Twitter and the Scottish Referendum Can Tell Us.* MSc Conflict Analysis and Resolution, George Mason University. Available: http://ebot.gmu.edu/bitstream/handle/1920/9767/Morris_thesis_2015.pdf?sequence=1&isAllowed=y [Accessed: 2017-07-14].

Morstatter, F., Pfeffer, J., Liu, H. and Carley, K.M. (2013) Is the Sample Good Enough? Comparing data from Twitter's Streaming API with Twitter's Firehose. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media.* Boston, Mass, July 8, 2013. Paolo Alto: Association for the Advancement of Artificial Intelligence , pp. 400.

Murakami Wood, D. and Webster, W. (2009) Living in Surveillance Societies: the Normalisation of Surveillance in Europe. *Journal of Contemporary European Research,* 5 (2), pp. 259-273.

Nakamura, R., & Smallwood, F. (1980). The Policy Implementation. New York: St. Martin's

Nakamura, R.T. (1987) 'The textbook policy process and implementation research'. *Review of Policy Research,* 7 (1), pp. 142-154.

Nasukawa, T. and Yi, J. (2003) Sentiment analysis: capturing favorability using natural language processing. *Proceedings of the 2nd International Conference on Knowledge Capture.* Sanibel Island, Florida, 23-25th October 2003. New York: ACM Publications , pp. 70-77. doi: 10.1145/945645.945658 [Accessed: 2017-06-23].

National Audit Office (1986) *Report by the Comptroller and Auditor General: The Rayner Scrutinies 1979-83.* Her Majesty's Stationery Office.

Neumayer, C., & Rossi, L. (2016). 15 Years of Protest and Media Technologies Scholarship: A Sociotechnical Timeline, *Social Media + Society* https://doi.org/10.1177/2056305116662180

Nissenbaum, H. (2011) A Contextual Approach to Privacy Online. *Daedelus,* 140 (4), pp. 32-48.

Norris, C. (2004) Opening Remarks. *CCTV and Social Control: The Politics and Practice of Video Surveillance Conference,* Centre for Criminological Research, University of Sheffield, 7-9 January

Office for National Statistics (2011), 2011 Census: Population Estimates by five-year age bands, and Household Estimates, for Local Authorities in the United Kingdom, Available: http://ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuspopulationestimatesbyfiveyearagebandsandhouseholdestimatesforlocalauthoritiesintheunitedkingdom [Accessed: 2019-07-02]

O'Malley, P., Weir, L. and Shearing, C. (1997) 'Governmentality, Criticism, Politics'. *Economy and Society.,* 26 (4), pp. 501-517.

Omand, D., Bartlett, J. and Miller, C. (2012) Introducing Social Media Intelligence (SOCMINT). *Intelligence and National Security.,* 27 (6), pp. 801-823.

O'Reilly, T. (2007) What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies,* (1), pp. 17.

Organisation for Economic Co-operation and Development (2003) *Managing Decentralisation: A New Role for Labour Market Policy.* OECD Publishing.

Ostrom, E. and Ostrom, V. (2011) The Institutional Perspective on Values and Virtues. In: M.S. de Vries and P. Kim, eds. *Value and Virtue in Public Administration A Comparative Perspective.* London: Palgrave, pp. 115-121.

Panger, G. (2016) Reassessing the Facebook experiment: critical thinking about the validity of Big Data research. *Information, Communication & Society,* 19 (8), pp. 1108-1126.

Parker, A.M., Bruine de Bruin, W. and Fischhoff, B. (2007) Maximizers versus satisficers: Decisionmaking styles, competence, and outcomes. *Judgement and Decision Making,* 2 (6), pp. 342-350.

Paudel, N.R. (2009) A Critical account of policy implementation theories. *Nepalese Journal of Public Policy and Governance,* XXV (2), pp. 36.

Porter, B. (1991). *The Origins of the Vigilant State: The London Metropolitan Police Special Branch before the First World War.* Boydell & Brewer.

Pressman, J.L. and Wildavsky, A. (1973) *Implementation : how great expectations in Washington are dashed in Oakland ... : this being a saga of the Economic Development Administration* . Berkeley: University of California Press.

Protevi, J. (2006) Deleuze, Guattari and Emergence. *A Journal of Modern Critical Theory,* 29 (2), pp. 19-39.

Protevi, J. (2012) Deleuze and Life. In: D.W. Smith and H. Somers-Hall, eds. *The Cambridge Companion to Deleuze.* 1st ed. Cambridge: Cambridge University Press, pp. 239-264.

Public Welfare/ Protection of Human Subjects 2009, Code of Federal Regulations 45 CFR 46. Available: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html, Available: Department of Health and Human Services [Accessed: 2016-02-02].

Renaud, K., Flowerday, S., English, R., & Volkamer, M. (2016). Why don't UK citizens protest against privacy-invading dragnet surveillance?. *Information & Computer Security.*

Robinson, N. (2003) Fuel Protests: Governing the Ungovernable? *Parliamentary Affairs,* 56 (3), pp. 423-440.

Rogowski, R. (2013) Shooting (or Ignoring) the Messenger. *Political Studies Review,* 11 (2), pp. 216-221.

Rose, N. (1999) *Powers of Freedom: Reframing Political Thought.* Cambridge: Cambridge University Press.

Rose, R. and Davies, P.L. (1994) *Inheritance in public policy : change without choice in Britain.* London: Yale University Press.

Rosenzweig, P. (2010) Privacy and Counter-Terrorism: The Pervasiveness of Data; Rosenzweig, Paul. *Case W. Res. J. Int'l L.,* 42, pp. 625.

Sabatier, P., 1987 Knowledge, Policy-oriented learning and policy change: An Advocacy Coalition Framework, *Knowledge,* 8(4), pp.649-692

Sayer, A. (1992) *Method in Social Science*, London: Routledge, 2nd edition

Schattschneider, E. E. (1975). *The semisovereign people: A realist's view of democracy in America*. Wadsworth Publishing Company.

Schneider, A. L., & Ingram, H. M. (1997). Policy design for democracy. University Press of Kansas.

Schütz, A., (1972) Trans Lehnert, F., Walsh, G. *The phenomenology of the social world; with an introduction by George Walsh*. London: Heinemann Educational.

Schwartz, M., Pre-Occupied: The origins and future of Occupy Wall Street, *The Political Scene,* New Yorker, 28th Nov 2011, Available: http://www.newyorker.com/reporting/2011/11/28/111128fa_fact_schwartz?currentPage=all [Accessed 2019-10-06]

Selznick, P. (1948) Foundations of the Theory of Organization. *American Sociological Review,* 13 (1), pp. 25-35.

Semetko, H.A. and Valkenburg, P.M.,. (2000) Framing European Politics: A Content Analysis of Press and Television news. *Journal of Communication,* 50 (20), pp. 93-109.

Semiocast (2010) *UK General Election 2010, 2nd Debate 22nd April : Semiocast Analysis of Comments and Reactions on Twitter.* Semiocast. Available: https://semiocast.com/downloads/Semiocast_twitter_analysis_uk_general_elections_debate_2_20100422.pdf [Accessed: 2015-08-01].

Sen, A. and Sen, A. (2000) The Discipline of Cost-Benefit Analysis. *The Journal of Legal Studies.,* 29 (2), pp. 931-952.

Simon, H.A. (1972) Theories of bounded rationality. *Decision and Organization,* 1, pp. 161-176.

Solberg, L.B. (2010)
Data Mining on Facebook: A Free Space for Researchers or an IRB Nightmare *University of Illinois. J.L. Tech. & Pol'y,* 311.

Srnicek, N. (2007) *Assemblage Theory, Complexity and Contentious Politics. The Political Ontology of Gilles Deleuze.* Masters, University College London.

Stake, R. (2005) *Multiple Case Study Analysis.* England: Guilford.

Staples, W.G. (1997) *The Culture of Surveillance: Discipline and Social Control in the United States. New York: St. Martin's Press.* New York: St Martin's Press.

Steinmo, S. (1993). *Taxation and democracy: Swedish, British, and American approaches to financing the modern state*. Yale University Press.

Stenson, K. (2005) 'Sovereignty, Biopolitics and the Local Government of Crime in Britain'. *Theoretical Criminology,* 9, pp. 265-287.

Stokey, E., & Zeckhauser, R. (1978). *Primer for policy analysis*. WW Norton.

The Verge (2017), *WhatsApp reportedly refused to build a backdoor for the UK government,* https://www.theverge.com/2017/9/20/16338128/whatsapp-reportedly-refused-request-uk-government-access-encrypted-messages [Accessed 08-10-2019]

Thissen, W.A.H. and Walker, W.E. eds. (2013) *Public Policy Analysis : New Developments.* New York: Springer US.

Thomas, G. (2011) A typology for the case study in social science following a review of definition, discourse and structure. *Qualitative Inquiry,* 17 (6), pp. 511-521.

Tokars, J.I., Richards, C., Andrus, M., Klevens, M., Curtis, A., Horan, T., Jernigan, J. and Cardo, D. (2004) The Changing Face of Surveillance for Health Care: Associated Infections. *Clinical Infectious Diseases,* 39 (9), pp. 1347-1352.

Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M. (2010) Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Fourth International Conference on Weblogs and Social Media.* Washington DC, 2010-05-23. , pp. 178. Available: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1441/1852 [Accessed: 2017-08-21].

Twitter (2014) *@TwitterSupport: You can now thread your tweets together.* Twitter.com: Twitter. Available: https://twitter.com/twittersupport/status/442433903546994688?lang=en [Accessed: 2017-12-15].

Twitter *Sample Real Time Tweets.* developer.twitter.com: Twitter. Available: https://developer.twitter.com/en/docs/tweets/sample-realtime/overview/decahose [Accessed: 2017-12-01].

Twitter (2007) *Are you twittering me?* blog.twitter.com: Twitter. Available: https://blog.twitter.com/official/en_us/a/2007/are-you-twittering-me.html [Accessed: 2017-12-05].

Twitter (2019), *Information Requests* https://transparency.twitter.com/en/information-requests.html [Accessed 01-11-2019]

Tversky, A. and Kahneman, D. (1986) Rational Choice and the Framing of Decisions. *The Journal of Business,* 59 (4, Part 2: The Behavioral Foundations of Economic Theory), pp. S251-S278.

Vickers, G. (1965) *The Art of Judgement: A Study Of Policymaking.* 1995 ed. London: Sage.

Walker, W.E., Rahman, S.A. and Cave, J. (2001) Adaptive policies, policy analysis, and policy-making. *European Journal of Operational Research,* 128, pp. 282.

Walker, C., & Hyland, K. (2014). *Undercover policing and underwhelming laws.* Available at SSRN 2469190.

Wall, D. (2007) *Cybercrime: The Transformation of Crime in the Information Age.* First Edition ed. Cambridge: Polity Press.

Wallsten, K. (2005) Political blogs and the bloggers who blog them: Is the political blogosphere and echo chamber. *American Political Science Association's Annual Meeting. Washington, DC, September*, pp. 1-4.

Webster, C. W. R. (2009). CCTV policy in the UK: Reconsidering the evidence base. *Surveillance and society*, 6(1), 10-22.

Whitaker, R. (1999) *The End of Privacy: How Total Surveillance is Becoming a Reality.* New York: The New Press.

Wildavsky, A. (1979) *Speaking Truth to Power: The Art and Craft of Policy Analysis.* Boston: Little Brown.

William, N. D. (1994). Public policy analysis: An introduction.

Wright, S. (2005) The ECHELON Trail: An illegal vision. *Surveillance and Society,* 3 (2/3), pp. 198.

Yin, R.K. (2003) *Case Study Research: Design and Methods.* 3rd ed. Thousand Oaks.

Yougov (2015) *General Election 2015: How Britain Really Voted.* London: Yougov. Available: https://d25d2506sfb94s.cloudfront.net/cumulus_uploads/document/owc2a5orpr/Establishment_Extra_Variables_Website.pdf [Accessed: 2017-04-11].

Appendix 1: Tweet Harvester Code

```
__author__ = 'Chris Campbell christopher.campbell@stir.ac.uk'
def read_user(tweet):
    from harvester.models import User
    from django.core.exceptions import ObjectDoesNotExist,
MultipleObjectsReturned
    #We might get results where user has changed their
details"], so first we check the UID.
    #print "MULTIPLE USER DEBUG", tweet["user"]["id_str"]
    try:
        current_user =
User.objects.get(id_str=tweet.user.id_str)
        created=False
        return current_user, created
    except ObjectDoesNotExist:
        pass
    except MultipleObjectsReturned:
        current_user =
User.objects.filter(id_str=tweet["user"]["id_str"])[0]
        return current_user, False
    if not tweet["user"]["follow_request_sent"]:
        tweet["user"]["follow_request_sent"] = False
    if not tweet["user"]["following"]:
        tweet["user"]["following"] = False
    if not tweet["user"]["description"]:
        tweet["user"]["description"] = " "
    if not tweet["user"]["notifications"]:
        tweet["user"]["notifications"] = False
    #If that doesn't work"],  use get_or_create (as a failback
rather than save())
    from dateutil.parser import parse
    if not tweet["user"]["contributors_enabled"]:
        current_user, created = User.objects.get_or_create(

follow_request_sent=tweet["user"]["follow_request_sent"],
            _json = {},
            verified = tweet["user"]["verified"],
            followers_count =
tweet["user"]["followers_count"],
            profile_image_url_https =
tweet["user"]["profile_image_url_https"],
            id_str = tweet["user"]["id_str"],
            listed_count = tweet["user"]["listed_count"],
            utc_offset = tweet["user"]["utc_offset"],
            statuses_count = tweet["user"]["statuses_count"],
            description = tweet["user"]["description"],
            friends_count = tweet["user"]["friends_count"],
```

```
            location = tweet["user"]["location"],
            profile_image_url=
tweet["user"]["profile_image_url"],
            following = tweet["user"]["following"],
            geo_enabled = tweet["user"]["geo_enabled"],
            profile_background_image_url
=tweet["user"]["profile_background_image_url"],
            screen_name = tweet["user"]["screen_name"],
            lang =  tweet["user"]["lang"],
            profile_background_tile =
tweet["user"]["profile_background_tile"],
            favourites_count =
tweet["user"]["favourites_count"],
            name = tweet["user"]["name"],
            notifications = tweet["user"]["notifications"],
            url = tweet["user"]["url"],
            created_at = parse(tweet["user"]["created_at"]),
            contributors_enabled = False,
            time_zone = tweet["user"]["time_zone"],
            protected = tweet["user"]["protected"],
            default_profile =
tweet["user"]["default_profile"],
            is_translator = tweet["user"]["is_translator"]
        )
    else:
        current_user, created = User.objects.get_or_create(

follow_request_sent=tweet["user"]["follow_request_sent"],
            _json = {},
            verified = tweet["user"]["verified"],
            followers_count =
tweet["user"]["followers_count"],
            profile_image_url_https =
tweet["user"]["profile_image_url_https"],
            id_str = tweet["user"]["id_str"],
            listed_count = tweet["user"]["listed_count"],
            utc_offset = tweet["user"]["utc_offset"],
            statuses_count = tweet["user"]["statuses_count"],
            description = tweet["user"]["description"],
            friends_count = tweet["user"]["friends_count"],
            location = tweet["user"]["location"],
            profile_image_url=
tweet["user"]["profile_image_url"],
            following = tweet["user"]["following"],
            geo_enabled = tweet["user"]["geo_enabled"],
            profile_background_image_url
=tweet["user"]["profile_background_image_url"],
            screen_name = tweet["user"]["screen_name"],
            lang =  tweet["user"]["lang"],
            profile_background_tile =
tweet["user"]["profile_background_tile"],
```

```
            favourites_count =
tweet["user"]["favourites_count"],
            name = tweet["user"]["name"],
            notifications = tweet["user"]["notifications"],
            url = tweet["user"]["url"],
            created_at = parse(tweet["user"]["created_at"]),
            contributors_enabled =
tweet["user"]["contributers_enabled"],
            time_zone = tweet["user"]["time_zone"],
            protected = tweet["user"]["protected"],
            default_profile =
tweet["user"]["default_profile"],
            is_translator = tweet["user"]["is_translator"]
        )
    #print "CURRENT USER:""], type(current_user)"],
current_user
    #current_user"], created =
User.objects.get_or_create(current_user)
    return current_user, created
def read_tweet(tweet, current_user):
    import logging
    logger = logging.getLogger('django')
    from datetime import date, datetime
    #print "Inside read_Tweet"
    from harvester.models import Tweet
    from django.core.exceptions import ObjectDoesNotExist,
MultipleObjectsReturned
    from django.db import DataError
    #We might get weird results where user has changed their
details"], so first we check the UID.
    #print tweet_data["created_at"]
    from dateutil.parser import parse
    tweet["created_at"] = parse(tweet["created_at"])
    try:
        #print "trying tweet_data["id"
        current_tweet
=Tweet.objects.get(id_str=tweet["id_str"])
        created=False
        return current_user, created
    except ObjectDoesNotExist:
        pass
    except MultipleObjectsReturned:
        current_tweet
=Tweet.objects.filter(id_str=tweet["id_str"])[0]
    try:
        current_tweet, created = Tweet.objects.get_or_create(
        truncated=tweet["truncated"],
        text=tweet["text"],
        favorite_count=tweet["favorite_count"],
        author = current_user,
        _json = tweet["_json"],
```

```
            source=tweet["source"],
            retweeted=tweet["retweeted"],
            coordinates = tweet["coordinates"],
            entities = tweet["entities"],
            in_reply_to_screen_name =
tweet["in_reply_to_screen_name"],
            id_str = tweet["id_str"],
            retweet_count = tweet["retweet_count"],
            favorited = tweet["favorited"],
            user = tweet["user"],
            geo = tweet["geo"],
            in_reply_to_user_id_str =
tweet["in_reply_to_user_id_str"],
            lang = tweet["lang"],
            created_at = tweet["created_at"],
            place = tweet["place"])
            print "DEBUG", current_user, current_tweet
            return current_tweet, created
        except DataError, e:
            #Catchall to pick up non-parsed tweets
            print "DEBUG ERROR", e, tweet
            return None, False
def read_both(tweet):
    current_user, created = read_user(tweet)
    current_tweet, created = read_tweet(tweet, current_user)
    print "USER", created, current_user
    print "TWEET", created, current_tweet
def queryset_iterator(queryset, chunksize=1000):
    import gc
    '''''
    Iterate over a Django Queryset ordered by the primary key
    This method loads a maximum of chunksize (default: 1000)
rows in it's
    memory at the same time while django normally would load
all rows in it's
    memory. Using the iterator() method only causes it to not
preload all the
    classes.
    Use the garbage collector library 'gc' to actively manage
the memory after each iteration
    and wipe previous queryset from memory.
    Note that the implementation of the iterator does not
support ordered query sets.
    '''
    pk = 0
    last_pk = queryset.order_by('-pk').first().pk
    queryset = queryset.order_by('pk')
    while pk < last_pk:
        print "pk", pk
        for row in queryset.filter(pk__gt=pk)[:chunksize]:
            pk = row.pk
```

```
            yield row
        gc.collect()
def add_string_codes():
    from harvester.models import Tweet
    replies =
Tweet.objects.exclude(in_reply_to_user_id_str__isnull=False)
    with open('id_lists.txt', 'ab+') as f:
        for tweet in queryset_iterator(replies):
            f.write(tweet.id_str+", ")

def chunker(seq, size):
    # Function to break a large list or query set into smaller
    # more manageable chunks.  Not memory efficient for
Querysets.
    return (seq[pos:pos + size] for pos in xrange(0, len(seq),
size))
```

Appendix 2: tools.py

```
__author__ = 'Chris Campbell christopher.campbell@stir.ac.uk'

"""Set of manual tools for data management and checking, and
data state change on the way into the database."""


def check_date(date_in):
    """Ensure dates are datetime objects rather than
strings"""
    from datetime import datetime
    if not isinstance(date_in, datetime):
        date_out = datetime.strptime(date_in,
"%Y-%m-%d-%H-%M-%S")
    else:
        date_out = date_in
    return date_out



def count_tweets(start, end):
    """Count the number of tweets in a specified period."""
    from harvester.models import Tweet
    start = check_date(start)
```

```
    end = check_date(end)

    number_tweets =
Tweet.objects.filter(created_at__gte=start,
created_at__lte=end, lang='en').count()

    return {"start": start, "end": end, "tweets":
number_tweets}




def tweets_over_time(tweet_set):

    """ Count the number of tweets per minute over a specified
     period"""

    from django.db.models.functions import Trunc

    from django.db.models import DateTimeField, Count

    tpm = tweet_set.annotate(

        start_min = Trunc('created_at', 'minute',

                          output_field=DateTimeField())).\

        values('start_min').\

        annotate(tweets=Count('pk'))

    return tpm




def tweets_over_time(start, end, period, tweets=None):
```

```
    “”” Count the number of tweets per timedelta over a
specified
    time period.  Can assess tweet queryset passed in, and if
no
    queryset is passed in, will count over all tweets.”””
    from harvester.models import Tweet
    from datetime import datetime, timedelta
    time_dict = {"seconds": 1,
             "tenminutes": 60*10,
             "thirtyseconds": 30,
             "minutes": 60,
             "hour":60*60,
             "day": 60*60*24}
    start = check_date(start)
    end = check_date(end)
    if not tweets:
        tweets = Tweet.objects.filter(created_at__gte=start,
created_at__lte=end, lang='en').__author__ = 'chris'


def clean_tweet_set(tweet_set=None):
    from harvester.models import Tweet
    from django.db.models import Q
```

```
    """
    Filters out a list of cruft either from an intake
tweet_set, or from 'all tweets'
    :return:
    """
    exclusion_list = ['cardiac arrest', 'CARD/RESP ARREST',
                      '#fit',
                      '#androidgames',
                      '#iphonegames',
                      'http://t.co/1aG5q2YMlw',
                      '#ipadgames', 'Good Gut Bacteria',
                      '#fatloss', '#instadeily',
                      'asaram bapu ji', 'Maher Arar', 'Michael
Barrymore',
                      'Delhi govt raided JIMS Jagran Institute if
management studies in Rithala',
                      'BJP', "Uganda's ex PM defiant after
arrest",
                      'Tomar', 'Nebraska Murder', ' shooting of
incoming BYU recruit',
                      '#aap', 'Nigerian Troops Arrest Jos',
'#GameInsight',
                      'Nigeria Arrest Former Chief Security
Officer',  '#healtylife',
```

```
                    '#mystyle', '#ootd', '#instagramers',

                    'Buhari reportedly planning to arrest',

                    'kpk', 'EFCC confirms arrest of former
governors Ohakim',

                    'Amama Mbabazi', 'JahangirKTareen', 'Godwin
Obua',

                    'Manoj Kumar', 'Muhammadu Buhari', 'Federal
Marshals Arrest 12-Year-Old, and Accomplices',

                    'Federal Agents Arrest Baby-Faced Boy',

                    'Taeba Darwish', '#jungkook', '#jin',
'#suga', 'Peoples Democratic Party',

                    'Xiaomi Mi5', 'Moses Kuria', 'Police arrest
masquerade in Enugu for stabbing priest',

                    'Layla Al Qaseer', 'Raihana Mosawi', 'Jalela
sayed ameen',

                    'Nigerian Troops Arrest Mastermind Of Jos',
'in Karachi operation'

                    '#FitnessFriday', '#girlswholift',
'#deadlift', '#fitfam', '#Palestinian', '#ModiNotWelcome']

    q_objects = Q()

    if not tweet_set:

        tweet_set = Tweet.objects.all()

    # loop trough the list and create an OR condition for each
item

    for item in exclusion_list:
```

```
        q_objects.add(Q(text__icontains=item), Q.OR)
        return tweet_set.exclude(q_objects)


def million_mask():
    """ Function to pull tweets for Million Mask march and
    clean them against the exclusion list. """
    from harvester.models import Tweet
    from django.db.models import Q
    from django.core.cache import cache
    #Get tweets we know are directly relevant from dataset
    relevant = Tweet.objects.filter(Q(text__icontains="Million
Mask March") |

Q(text__icontains="millionmaskmarch") |

Q(text__icontains="mmm2015"))
    print "Core tweets pulled"
    trimmed = clean_tweet_set(relevant)
    relevant = None # explicit clear for garbage collection!
    print "Core tweets cleaned"
    return trimmed
```

```
def pull_out_links(tweet):
    tokens = tweet.text.split()
    links = [token for token in tokens if 'http' in token]
    return (links, tweet.id)



def tweets_over_time_graph(data):
    import pygal, datetime
    new = [(key, float(value)) for (key, value) in
data['tweets'].items()]
    line_chart = pygal.Bar()
    x_axis = [item for item in sorted(data['tweets'])]
    series = [data['tweets'][item] for item in
sorted(data['tweets'])]
    line_chart.x_labels = x_axis
    line_chart.add('Tweets', series)
    line_chart.render_to_file('')
def count_period(start, end, filtered):
    from django.core.cache import cache
    from harvester.models import Tweet
```

```
    if filtered:
        tweets = Tweet.objects.filter(created_at__gte=start,
created_at__lte=end)
        results = clean_tweet_set(tweets).count()
    else:
        results = Tweet.objects.filter(created_at__gte=start,

created_at__lte=end).count()
    saved_results = cache.get('filtered_by_hour')
    if not saved_results:
        saved_results = {}
    saved_results[start] = results
    cache.set('filtered_by_hour', saved_results)
    return



def tweets_by_hour(start, end, filtered=False):
    """
    Function that only pulls tweets by hour, instead of
pulling them all and filtering by hour, which has memory
    capacity and speed problems
    :param start:
```

```
    :param end:

    :return: data series dictionary in {'start_time':
YYYYMMDD:H:M:S:m, "number_of_tweets"} format

    """

    from datetime import timedelta

    from django_rq import enqueue

    working_end= start + timedelta(hours=1)

    working_start = start

    results = {}

    while working_start < end:

        print "Working time", working_start

        enqueue(count_period, working_start, working_end,
filtered)

        working_end = working_end + timedelta(hours=1)

        working_start = working_start + timedelta(hours=1)

    return

def test_political_presence():

    from harvester.models import User

    from django.db.models import Q

    """

    Finds and filters users by the presence of political party
mentions in their user description field
```

```
    """
    search_list = ['Conservative', 'Labour', 'Lib Dem',
                   'Liberal Democrat', 'SNP', 'UKIP', 'UK
independence Party', 'Green Party']
    q_objects = Q()
    exclude_qs = Q()
    exclude_list = ['Republican', 'Democrat']
    users = User.objects.all()
    # loop trough the list and create an OR condition for each
item
    for item in search_list:
        q_objects.add(Q(text__contains=item), Q.OR)
    for item in exclude_list:
        exclude_qs.add(Q(text__contains=item), Q.OR)
    pols = users.filter(q_objects).exclude(exclude_qs)
    return pols
def display_users_by_tweets():
    from django.db.models import Count
    from harvester.models import User
    large_users =
User.objects.all().annotate(num_tweets=Count('tweet')).order_b
y('-num_tweets')
```

```
def june_20th_tweets():
    """
    Pulls the tweets for June twentieth protest
    :return:
    """
    from datetime import datetime
    from harvester.models import Tweet
    from django.db.models import Q
    start = datetime(2015,6,18,3,0)
    end = datetime(2015,6,22,19,0)
    initial = Tweet.objects.filter(created_at__gte=start,
                                   created_at__lte=end)
    content = initial.filter(
    Q(text__icontains='protest') |
    Q(text__icontains='austerity') |
    Q(text__icontains='demonstrat') |
    Q(text__icontains='poor') |
    Q(text__icontains='police') |
    Q(text__icontains='surveil')
)
```

```
print "Initialtweets pulled June 20th"

cleaned = clean_tweet_set(initial)

initial = None #explicit empty for garbage collection

print "Initial tweets cleaned and cleared"

return cleaned, content
```

Appendix 3: Latent Semantic Analysis Code

```
from numpy import zeros, transpose, asarray, sum, diag, dot,
arccos

from numpy.linalg import norm

import numpy

from scipy.linalg import svd, inv

import matplotlib.pyplot as plt

import re, random, pylab

from math import *

from operator import itemgetter

# Adapted from https://github.com/kernelmachine/pyLSA

# stopwords, retreived from
http://www.lextek.com/manuals/onix/stopwords1.html

stopwords = ["i'm", "#endausteritynow", "&amp;amp;", "@", "don't",
"\u2026", "rt", "you're", "can't", "didn't"

            'a', 'about', 'above', 'across', 'after',
'again', 'against',

            'all', 'almost', 'alone', 'along', 'already',
'also', 'although', 'always',

            'among', 'an', 'and', 'another', 'any',
'anybody', 'anyone', 'anything',
```

```
            'anywhere', 'are', 'area', 'areas', 'around',
'as', 'ask', 'asked',

            'asking', 'asks', 'at', 'away', 'b', 'back',
'backed', 'backing', 'backs', 'be',

            'became', 'because', 'become', 'becomes', 'been',
'before', 'began', 'behind',

            'being', 'beings', 'best', 'better', 'between',
'big', 'both', 'but', 'by', 'c',

            'came', 'can', 'cannot', 'case', 'cases',
'certain', 'certainly', 'clear', 'clearly',

            'come', 'could', 'd', 'did', 'differ',
'different', 'differently', 'do', 'does', 'done',

            'down', 'down', 'downed', 'downing', 'downs',
'during', 'e', 'each', 'early', 'either',

            'end', 'ended', 'ending', 'ends', 'enough',
'even', 'evenly', 'ever', 'every', 'everybody',

            'everyone', 'everything', 'everywhere', 'f',
'face', 'faces', 'fact', 'facts', 'far',

            'felt', 'few', 'find', 'finds', 'first', 'for',
'four', 'from', 'full', 'fully',

            'further', 'furthered', 'furthering', 'furthers',
'g', 'gave', 'general', 'generally',

            'get', 'gets', 'give', 'given', 'gives', 'go',
'going', 'good', 'goods', 'got', 'great',

            'greater', 'greatest', 'group', 'grouped',
'grouping', 'groups', 'h', 'had', 'has', 'have',
```

```
            'having', 'he', 'her', 'here', 'herself', 'high',
'high', 'high', 'higher', 'highest',

            'him', 'himself', 'his', 'how', 'however', 'i',
'if', 'important', 'in', 'interest',

            'interested', 'interesting', 'interests', 'into',
'is', 'it', 'its', 'itself', 'j',

            'just', 'k', 'keep', 'keeps', 'kind', 'knew',
'know', 'known', 'knows', 'l', 'large', 'largely',

            'last', 'later', 'latest', 'least', 'less',
'let', 'lets', 'like', 'likely', 'long', 'longer',

            'longest', 'm', 'made', 'make', 'making', 'man',
'many', 'may', 'me', 'member', 'members',

            'men', 'might', 'more', 'most', 'mostly', 'mr',
'mrs', 'much', 'must', 'my', 'myself',

            'n', 'necessary', 'need', 'needed', 'needing',
'needs', 'never', 'new', 'new', 'newer',

            'newest', 'next', 'no', 'nobody', 'non', 'noone',
'not', 'nothing', 'now', 'nowhere',

            'number', 'numbers', 'o', 'of', 'off', 'often',
'old', 'older', 'oldest', 'on',

            'once', 'one', 'only', 'open', 'opened',
'opening', 'opens', 'or', 'order',

            'ordered', 'ordering', 'orders', 'other',
'others', 'our', 'out', 'over', 'p',

            'part', 'parted', 'parting', 'parts', 'per',
'perhaps', 'place', 'places', 'point',
```

```
            'pointed', 'pointing', 'points', 'possible',
'present', 'presented', 'presenting',

            'presents', 'problem', 'problems', 'put', 'puts',
'q', 'quite', 'r', 'rather',

            'really', 'right', 'right', 'room', 'rooms', 's',
'said', 'same', 'saw', 'say',

            'says', 'second', 'seconds', 'see', 'seem',
'seemed', 'seeming', 'seems',

            'sees', 'several', 'shall', 'she', 'should',
'show', 'showed', 'showing',

            'shows', 'side', 'sides', 'since', 'small',
'smaller', 'smallest', 'so',

            'some', 'somebody', 'someone', 'something',
'somewhere', 'state', 'states',

            'still', 'still', 'such', 'sure', 't', 'take',
'taken', 'than', 'that', 'the',

            'their', 'them', 'then', 'there', 'therefore',
'these', 'they', 'thing', 'things',

            'think', 'thinks', 'this', 'those', 'though',
'thought', 'thoughts', 'three',

            'through', 'thus', 'to', 'today', 'together',
'too', 'took', 'toward', 'turn',

            'turned', 'turning', 'turns', 'two', 'u',
'under', 'until', 'up', 'upon',

            'us', 'use', 'used', 'uses', 'v', 'very', 'w',
'want', 'wanted', 'wanting',
```

```
            'wants', 'was', 'way', 'ways', 'we', 'well',
'wells', 'went', 'were',

            'what', 'when', 'where', 'whether', 'which',
'while', 'who', 'whole', 'whose',

            'why', 'will', 'with', 'within', 'without',
'work', 'worked', 'working', 'works',

            'would', 'x', 'y', 'year', 'years', 'yet', 'you',
'young', 'younger', 'youngest', 'your',

            'yours', 'z']

ignore_characters = ''',:'!'''

def compare(queries):  # core comparison function.

    """ Compare two (or more) corpuses for similarity"""

    lsa = LSA(stopwords, ignore_characters)

    for q in queries:

        lsa.parse(q)

    lsa.build()

    lsa.calc()

    Vt = lsa.Vt

    S = diag(lsa.S)

    vectors = [(dot(S, Vt[:, 0]), dot(S, Vt[:, i])) for i in
range(len(Vt))]

    angles = [arccos(dot(a, b) / (norm(a, 2) * norm(b, 2)))
for a, b in vectors[1:]]

    return str(abs(1 - float(angles[0]) / float(pi / 2)))
```

```
def graph(query1, query2):
    """
    Kept but not used; plots the SVD for two queries in a
matplotlib graph.
    :param query1:
    :param query2:
    :return:
    """
    lsa = LSA(stopwords, ignore_characters)
    titles = [lsa.search_wiki(query1),
lsa.search_wiki(query2)]
    for t in titles:
        lsa.parse(t)
    lsa.build()
    lsa.calc()
    lsa.plotSVD()
## core summarization function.
def summarize(query=None, k=4, url=None):
    """
    Summarizes a given longer article
    :param query:
    :param k:
```

```
    :param url:

    :return:

    """

    j = []

    if url:

        b = URL(url)

        a = Document(b.download(cached=True))

        for b in a.get_elements_by_tagname("p"):

            j.append(plaintext(b.content).encode("utf-8"))

        j = [word for sentence in j for word in
sentence.split() if

             re.match("^[a-zA-Z_-]*$", word) or '.' in word or
"'" in word or '"' in word]

        j = ' '.join(j)

        lsa1 = LSA(stopwords, ignore_characters)

        sentences = j.split('.')

        sentences = [sentence for sentence in sentences if
len(sentence) > 1 and sentence != '']

        for sentence in sentences:

            lsa1.parse(sentence)

    else:

        lsa1 = LSA(stopwords, ignore_characters)
```

```
        sentences = query.split('.')
        for sentence in sentences:
            lsa1.parse(sentence)
    lsa1.build()
    lsa1.calc()
    summary = [(sentences[i], norm(dot(diag(lsa1.S),
lsa1.Vt[:, b]), 2)) for i in range(len(sentences)) for b in
                range(len(lsa1.Vt))]
    sorted(summary, key=itemgetter(1))
    summary = dict((v[0], v) for v in sorted(summary,
key=lambda summary: summary[1])).values()
    return '.'.join([a for a, b in summary][len(summary) -
(k):])

## evaluate the summarization. How well does the given summary
summarize the query?
def summarize_evaluation(query=None, url=None, summary=None):
    j = []
    if url:
        b = URL(url)
        a = Document(b.download(cached=True))
        for b in a.get_elements_by_tagname("p"):
            j.append(plaintext(b.content).encode("utf-8"))
```

```
        j = [word for sentence in j for word in
sentence.split() if
             re.match("^[a-zA-Z_-]*$", word) or '.' in word or
"'" in word or '"' in word]
        j = ' '.join(j)
        lsa = LSA(stopwords, ignore_characters)
        sentences = j.split('.')
        sentences = [sentence for sentence in sentences if
len(sentence) > 1 and sentence != '']
        for sentence in sentences:
            lsa.parse(sentence)
    else:
        lsa = LSA(stopwords, ignore_characters)
        for sentence in query:
            lsa.parse(sentence)
    lsa.build()
    lsa.calc()
    lsa2 = LSA(stopwords, ignore_characters)
    for sentence in summary:
        lsa2.parse(sentence)
    lsa2.build()
    lsa2.calc()
```

```
    vectors = [(dot(lsa.S, lsa.U[0, :]), dot(lsa.S, lsa.U[i,
:])) for i in range(len(lsa.U))]

    vectors2 = [(dot(lsa2.S, lsa2.U[0, :]), dot(lsa2.S,
lsa2.U[i, :])) for i in range(len(lsa2.U))]

    angles = [arccos(dot(a, b) / (norm(a, 2) * norm(b, 2)))
for a in vectors for b in vectors2]

    return str(abs(1 - float(angles[1]) / float(pi / 2)))

class LSA(object):

    def __init__(self, stopwords, ignore_characters):

        self.stopwords = stopwords

        self.ignore_characters = ignore_characters

        self.wdict = {}

        self.dcount = 0

    def parse(self, doc):

        if not isinstance(doc, list):

            words = doc.split()

        else:

            words = doc # already a list of words from the
corpus generator

        for w in words:

            w = w.lower()

            if w in self.stopwords:

                continue
```

```
            elif w in self.wdict:
                self.wdict[w].append(self.dcount)
            else:
                self.wdict[w] = [self.dcount]
        self.dcount += 1
    def build(self):  # Create count matrix
        self.keys = [k for k in self.wdict.keys() if
len(self.wdict[k]) > 1]
        self.keys.sort()
        self.A = zeros([len(self.keys), self.dcount])
        for i, k in enumerate(self.keys):
            for d in self.wdict[k]:
                self.A[i, d] += 1
    def calc(self):  # execute SVD
        self.U, self.S, self.Vt = svd(self.A,
full_matrices=False)
    def TFIDF(self):  # calculate tfidf score
        WordsPerDoc = sum(self.A, axis=0)
        DocsPerWord = sum(asarray(self.A > 0, 'i'), axis=1)
        rows, cols = self.A.shape
        for i in range(rows):
            for j in range(cols):
```

```
                self.A[i, j] = (self.A[i, j] / WordsPerDoc[j])
* log(float(cols) / DocsPerWord[i])

    def S(self):

        return self.S

    def U(self):

        return -1 * self.U

    def Vt(self):

        return -1 * self.Vt

    def printSVD(self):

        print 'Singular values: '

        print self.S

        print 'U matrix: '

        print -1 * self.U[:, 0:3]

        print 'Vt matrix: '

        print -1 * self.Vt[0:3, :]

    def search_wiki(self, k):  # scrape query's wikipedia
article

        article = Wikipedia().search(k)

        contents = [section.content.encode("utf8") for section
in article.sections]

        d = []

        for content in contents:
```

```
            a = content.split()
            d.append(a)
        content = [j for i in d for j in i if
                   re.match("^[a-zA-Z_-]*$", j) and len(j) >
1]  # take only meaningful content
        self.content = ' '.join(content)
        return self.content
    def plotSVD(self,
                k=5):  # change k to change how many points
you want to see on the graph. plots term vectors vs. document
vectors.
        y = numpy.random.random(10)
        d = numpy.random.random(10)
        fig = plt.figure()
        graph = fig.add_subplot(111)
        graph.autoscale(True)
        coordinates = [(s, a) for [s, a] in (-1 * self.U[:,
0:3]).tolist()]
        plot_coordinates = []
        for i in range(k):
            index = random.randint(1, len(coordinates))
            plot_coordinates.append(coordinates[index])
        xdata = [s for s, a in plot_coordinates]
```

```
        ydata = [a for s, a in plot_coordinates]

        plt.Arrow(0, 0, xdata[0], ydata[0])

        graph.scatter(xdata, ydata, c=y, s=20)

        graph.scatter(self.Vt[0:2, :].tolist()[0],
self.Vt[0:2, :].tolist()[1], marker='^', c=d, s=100)

        plt.show()

def create_corpus(tweet_queryset):

    """

    For a given queryset, return a sanitised-stopwords corpus.
Clears out stopwords and short length < 3 character

    words.  Does *not* filter content, see analysis.tools.py
for that.

    :param tweet_queryset:

    :return:

    """

    try:

        high_level_list = [tweet.text for tweet in
tweet_queryset]

    except AttributeError: # handle case where we only pull
textfield

        high_level_list = [tweet['text'] for tweet in
tweet_queryset]

    high_level_string = ' '.join(high_level_list)

    #convert to a tokenized list and strip stopwords
```

```
    high_level_string.lower().split()

    return [item for item in high_level_string.lower().split()
if item not in stopwords]

def run_comparison(start, end, timed_amount):

    from datetime import timedelta

    from harvester.models import Tweet

    from django.core.cache import cache

    corp = Tweet.objects.filter(created_at__gte=start,
created_at__lte=end)

    corp2 =
Tweet.objects.filter(created_at__gte=corp.last().created_at,


created_at__lte=corp.last().created_at + timedelta(hours=1))

    sanitised_1 = create_corpus(corp)

    sanitised_2 = create_corpus(corp2)

    res = compare([sanitised_1, sanitised_2])

    cache_dict = cache.get('results_dict')

    if not cache_dict:

        cache_dict = {}

    cache_dict[str(end)] = res

    print cache_dict

    cache.set('results_dict', cache_dict, None)

def compare_from_start(timed_amount=1):
```

```
    """
    Function mostly to capture the work that's been done here.
    :return:
    """
    from datetime import timedelta
    from harvester.models import Tweet
    from django_rq import enqueue
    t = Tweet.objects.all().first()
    start = t.created_at
    final = Tweet.objects.all().last().created_at
    end = t.created_at + timedelta(hours=timed_amount)
    while end < final:
        enqueue(run_comparison, start=start, end=end,
timed_amount=timed_amount)
        end = end + timedelta(hours=timed_amount)
        start = start + timedelta(hours=timed_amount)
    return
def rolling_tweet_analysis(tweet, offset_time=15,
surrounding_tweets=None):
    """
    Analyses a tweet for LSA to the preceeding and following X
minutes as a corpus.
```

```
    """
    from django.core.cache import cache
    from django.db.models import Q
    from harvester.models import Tweet
    from datetime import timedelta
    if isinstance(tweet, Tweet):
        pass
    else:
        tweet = Tweet.objects.get(id_str=tweet)
    if not surrounding_tweets:
        surrounding_tweets =
Tweet.objects.filter(Q(created_at__gte=tweet.created_at -
timedelta(minutes=offset_time),

created_at__lt=tweet.created_at) |

Q(created_at__lte=tweet.created_at +
timedelta(minutes=offset_time),

created_at__gt=tweet.created_at)).values('text')
    else:
        surrounding_tweets =
surrounding_tweets.filter(Q(created_at__gte=tweet.created_at -
timedelta(minutes=offset_time),
```

```
created_at__lt=tweet.created_at) |

Q(created_at__lte=tweet.created_at +
timedelta(minutes=offset_time),

created_at__gt=tweet.created_at)).values('text')
    results = cache.get('rolling_comparison')
    if not results:
        results = {}
    sanitised_surrounding = create_corpus(surrounding_tweets)
    sanitised_tweet = create_corpus([tweet])
    #append the output of that to the dictionary from cache
and re set the cache.
    results[tweet.id_str] = {'lsa':
compare([sanitised_surrounding, sanitised_tweet]),
'created_at': tweet.created_at}
    cache.set('rolling_comparison', results, None)
    return
def rolling_mask_original():
    from .tools import million_mask
    from harvester.models import Tweet
    from django_rq import enqueue
    import time
```

```
    from datetime import timedelta
    core_tweets = million_mask().order_by('created_at')
    print "core and additional pulled, queuing %d tweets" %
(core_tweets.count(),)
    #start = core_tweets.first().created_at
    #end = core_tweets.last().created_at
    for tweet in core_tweets:
        enqueue(rolling_tweet_analysis, tweet)
def rolling_ten_minute_window(selection='mask'):
    from .tools import million_mask, june_20th_tweets
    from harvester.models import Tweet
    from datetime import timedelta
    from django.core.cache import cache
    if selection == 'mask':
        core_tweets = million_mask().order_by('created_at')
    elif selection == 'june':
        core_tweets =
june_20th_tweets().order_by('created_at')
    else:
        raise NotImplementedError("Mo such selection")
    print "core and additional pulled, analysing %s" %
(core_tweets.count())
```

```
    start = core_tweets.first().created_at

    end = core_tweets.last().created_at

    print "Start: %s End: %s" % (str(start), str(end))

    working_time_end = start + timedelta(minutes=10)

    working_time_start = start

    while working_time_end < end:

        results = cache.get('maskals')

        if not results: results = {}

        current_tweets =
core_tweets.filter(created_at__gte=working_time_start,


created_at__lte=working_time_end)

        print "Current", current_tweets

        window_start =
working_time_start-timedelta(minutes=30)

        window_end = working_time_end+timedelta(minutes=30)

        surrounding_tweets =
Tweet.objects.filter(created_at__gte=window_start,


created_at__lte=window_end).exclude(pk__in=current_tweets)

        print "Surrounding", surrounding_tweets

        sanitised_surrounding =
create_corpus(surrounding_tweets)

        sanitised_current = create_corpus(current_tweets)
```

```
        print "Analysing %s" % (window_start,)
        #try:
        results[window_start] =
compare([sanitised_surrounding, sanitised_current])
        #except ValueError:
        #    results[window_start] = 'nan'
        cache.set('maskals', results, None)
        working_time_start = working_time_end
        working_time_end = working_time_end +
timedelta(minutes=10)
def generate_graph(stats_input, filename=None):
    import pygal
    if not filename:
        filename = "%s.svg" %
(stats_input.items()[0].strftime('%d, %b %Y at %I:%M:%S %p'))
    datetimeline = pygal.DateTimeLine(x_label_rotation=35,
                                      truncate_label=-1,
                                      x_value_formatter=lambda
dt: dt.strftime('%d, %b %Y at %I:%M:%S %p'))
    stats2 = [(key, float(value)) for key, value in
stats_input.items()]
    datetimeline.add('LSA output', stats2)
    datetimeline.render_to_file(filename)
```

```
    x_axis = [item for item in sorted(data['tweets'])]

    series = [data['tweets'][item] for item in
sorted(data['tweets'])]

    for thing in sorted(data['tweets']): print thing,
data['tweets'][thing]
```

Appendix 4: Sentiment Analysis Training Set Code

```
from textblob import TextBlob
import nltk
def get_tweet_sentiment(tweet):
    '''
    Utility function to classify sentiment of passed tweet
    using textblob's sentiment method
    '''
    # create TextBlob object of passed tweet text
    analysis = TextBlob(tweet.text)
    # set sentiment
    print analysis
    if analysis.sentiment.polarity > 0:
        return 'positive'
    elif analysis.sentiment.polarity == 0:
        return 'neutral'
    else:
        return 'negative'
from django.core.cache import cache
```

```
def train_classifier(tweet):
    """
    Used to build a training set of classified tweets for
sentiment; Asks the user to read the tweet and assess
    whether it is negative, positive, neutral, or should be
excluded as irrelevant.
    :param tweet:
    :return:
    """
    print tweet.text
    sentiment = raw_input("(n)egative, (p)ositive, (ne)utral
or (e)xclude.")
    options_dictionary = {
        'n': 'negative',
        'ne': 'neutral',
        'p': 'positive',
        'e': 'exclude',
    }
    return tweet.text, (options_dictionary[sentiment])
def build_excluded_keywords_filter():
    """
    Early attempt to build a keyword filter to reduce the size
of the working tweet set and return it as an ID
```

```
    list that can then be retrieved.  Has same problem as the
other attempts at this, of loading the oversized

    tweet objects (including 3x JSON representations) into
memory.

    :return:

    """

    from harvester.models import Tweet

    exclusion_keywords = ['delhi', 'bjp', 'tomar',
'arvindmantri']

    exclusion_list = []

    for item in exclusion_keywords:

        t = Tweet.objects.filter(text__icontains=item)

        exclusion_list.append(t)

        return  [item.id for sublist in exclusion_list for
item in sublist]

def random_tweets(number_of_tweets):

    """

    Select number_of_tweets randomly from the whole database.
Use to provide tweets for training classifier.

    :param number_of_tweets:

    :return:

    """

    from harvester.models import Tweet
```

```
    import random
    number = Tweet.objects.all().count()
    return random.sample(range(1, number), number_of_tweets)
def recursively_train():
    """
    Successively filter the queryset, trying to build up a
data set of 100 trained tweets as an initial setup.
    Each time we move through the tweet set we classify it as
above, or exclude it.  Next loop through the
    queryset filters out exclusion set again.
    :return:
    """
    from harvester.models import Tweet
    exclusion_keywords = ['delhi', 'bjp', 'tomar',
'arvindmantri']
    classified = 0
    tweets = Tweet.objects.filter(ids__in=random_tweets(5000))
    results = []
    exclude_text_filter = []
    while classified < 100:
        #after each classification/exclusion, add that text to
the filter, to ignore RTs, etc.
```

```
        current_tweets =
tweets.exclude(text__in=exclude_text_filter).order_by('created
_at')

        tweet = current_tweets.first() #next chronologically

        #because this is slow, we're just going to check it
and skip it in python, so oversample on the top end.

        if any(item in tweet.text for item in
exclusion_keywords):

            continue

        while True:

            try:

                tweettext, classification =
train_classifier(tweet)

            except KeyError:

                "Please enter a valid entry."

                continue

            break

        results.append({'tweet': tweettext, 'classification':
classification})

        cache.set('results', results, None) #save to memory so
we can access outside

        # print results

        exclude_text_filter.append(tweettext)
```

```
        #we should only add ones that are being actually
classified, rather than those we're excluding.
        if classification in ['negative', 'positive',
'neutral']:
            classified +=1
        print classified
    return results
def clean_tweets():
    #drop words two characters or less
    from django.core.cache import cache
    tweets = cache.get('classifier_training_set')
    output_set = []
    for (words, sentiment) in tweets:
        words_filtered = [e.lower() for e in words.split() if
len(e) >= 3]
        output_set.append((words_filtered, sentiment))
    return output_set
def get_words_in_tweets(tweets):
    """
    Create a corpus word-list from the list of tweets and
sentiments.
    :param tweets:
    :return:
```

```
    """
    all_words = []
    for (words, sentiment) in tweets:
      all_words.extend(words)
    return all_words
def get_word_features(wordlist):
    """
    Tags the corpus of words so that we can examine the most
common ones.
    :param wordlist:
    :return:
    """
    import nltk
    wordlist = nltk.FreqDist(wordlist)
    word_features = wordlist.keys()
    return word_features
def extract_features(document):
    document_words = set(document)
    features = {}
    for word in word_features:
        features['contains(%s)' % word] = (word in
document_words)
```

```
    return features
```

Appendix 5: Timeline of notable protests tied to social media

April 2009 - G20 protests in London, including the City of London, the Excel Centre in East London, and a Stop the War march through central London

May 2010 - Conservative/Liberal Democrat coalition government elected in UK

Nov-Dec 2010 - UK Student protests against tuition fees; Anti-Austerity protests in the UK; Conservative Central HQ at Millbank occupied and damaged;

December 2010 - Tunisian protests

January - March 2011 - Widespread protests and civil uprisings in multiple Arab countries including Oman, Yemen, Egypt, Syria and Morocco

March 2011 - Libyan protest movement escalates into civil uprising and civil war; GMB and UK Uncut protests in London and numerous UK cities; 500,000 attend TUC anti-austerity protest on 26th March

September 2011 - Occupy Wall Street campaign begins

October 2011 - Libyan Civil War officially ends, Occupy London campaign begins

August 2011 - London and wider England riots

November 2011 - Occupy Wall Street camp cleared forcefully from Zucotti Park in New York

January 2012 - Egyptian state of emergency, in place since 1967, lifted; Occupy London camp cleared from UBS and St Pauls

July 2012 - Syrian conflicts escalate into open warfare

October 2012 - TUC Marches against austerity in London, Belfast and Glasgow

January 2013 - Large scale protests in Egypt's Tahrir square, hundreds injured nationally

July 2013 - following protests and clashes with police and national security forces, Egypt's President Mohammed Morsi deposed in a coup

June 2014 - People's Assembly demonstration in London

September 2014 - Hong Kong protests 'Umbrella Revolution' begin

November 2014 - Million Mask March, further student protests in London

November 2015 - Student protests in London