DIGITAL BLACK BOXES: APPREHENDING FAST AND DYNAMIC SOCIOTECHNICAL NETWORKS

Danny Lee Weston

Thesis submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy awarded by the University of Greenwich

June 2016

DECLARATION

I certify that this work has not been accepted in substance for any degree, and is not concurrently being submitted for any degree other than that of a Ph.D. being studied at the University of Greenwich. I also declare that this work is the result of my own investigations except where otherwise identified by references and that I have not plagiarised another's work.

ACKNOWLEDGEMENTS

fought with (in or out of the ring) and whether we liked eachother or not. You all contributed to making me the unreasonable and awkward sod I am today, giving me just enough stubbornness to get through to the end. All of the sceptics who I must not name (you know who you are and no I'm not looking forward to your forensic questions if and when you read this…). All of the dearly missed absent friends, family and teachers (and Patrick, if you're actually still alive and ever read this, not a day has gone by without me thinking about you, or being inspired by your courage. You may not yet be the last PhD holder to turn to mercenary work nor to remember what John Adams said about studying war and politics).

And of course, finally, whisky and wine.

For Lucy,

and all the unspeakable paths you have had to travel.

ABSTRACT

We live in a world now dominated by complex and fast moving sociotechnical systems. This work considers the difficulties presented in even discussing such systems in an intelligible manner, especially given that many escape the immediate capabilities of human cognition to properly apprehend. Numerous means have in recent history been used to 'fix' or 'stabilise' the meanings and capabilities of such systems through one form of 'Black Box' or other. This work argues that our current conceptual resources are not appropriate to the task and explores ways in which both academics and practitioners can effectively and usefully apprehend such fast moving sociotechnical systems. It draws in particular on the work of Bruno Latour and Paul Thagard, combining their notions of 'Circulating Reference' and 'Computational Coherence' respectively, to propose more dynamic approaches and methodologies situated in kind of Computationalism that can be adapted to enable complex computing and digital networked events and the dynamic sociotechnical networks that underpin them to be talked about, and thereby *known* in ways that both express and encompass their dynamism and complexity and without inappropriately and prematurely placing them into 'Black Boxes'.

Contents

Figures

Chapter 1: Introduction

Digital systems and networks come in a wide variety of configurations. Many of these are simple and relatively self-contained. Others are so complex that they are beyond immediate human comprehension. This thesis concerns how the latter can be effectively apprehended, taking into account the speed and the complexity of the software agents and elements at work in current digital networks and the fact that they exist largely in the form of 'Black Boxes' to be opened and examined. I focus in particular here on how some computing technologies themselves can be effectively enrolled in such work.

In related work, Dodge and Kitchin explored how software is related to the creation, destruction and maintenance of space. They argue "the production of space is increasingly dependent on code, and code is written to produce space" (Dodge and Kitchin, 2011, preface). Following a corollary path of investigation, I explore here how the production of *narratives* is dependent on code and how code is written to produce narratives, especially when situated in fast-moving networks that overtake unassisted human perceptual and cognitive faculties. Specifically, I am looking at the multiple ontologies and epistemologies at work – both of which are effectively interrelated narratives about reality and in this case the reality of software and networks. Moreover, these narratives describe both events in the world and the complex software systems and networks themselves in a somewhat recursive fashion. For example, events recorded using digital equipment, transmitted across the world via digital networks, then shared and discussed on digital social media carry implicit narratives about these technologies themselves as well as the original event. Rogers, (2006), for example, demonstrated such phenomena actively at work in the use and application of search engines.

The level of complexity of such software is further deepened by an increasingly networked digital infrastructure – forming many webs, connections (not to mention multiple and often unpredictable inputs) that may only be partially visible and often beyond the tolerances anticipated by the original programmers. Machine Learning, Natural Language Processing and other techniques add further capacities, including the ability to adapt and self-modify. Yet these exist embedded within a social world of users, constrained and circumscribed by human behaviours and predilections, often (though not always) going completely awry when user attention and intervention is either completely absent or ineffective.

In such complex networks of actors, sometimes it is not clear who the human agents are and what is or is not the direct result of human action. Attributing responsibility in the face of such ambiguity then becomes extremely difficult. When considering the realm of digital communications for example, it is a relatively straightforward manner for humans and bots to masquerade as one another. "On the Internet, nobody knows you are a Bot," said David Kushner in *Wired* (Kushner, 2005). If one does not know who or what is or is not a 'bot', or what outcomes are the result of 'bot' or human action, where does one even start in attempting to describe, or map, what happens in any particular online or digital web of interactions? And what if, in fact, humans and machines are fundamentally interpenetrated in these events and actions – each enrolling the other continually so any such distinction is potentially meaningless?

Because of the pace such systems and networks operate at, analysis often takes place at a comparatively slow speed that necessarily looks backwards, generally proceeding in a post hoc manner. The human observer is often, in very real terms, already overtaken by events – not to mention by a compound of processes, bots and networks. The question is then how to describe such states of affairs and the interactions of their component parts in a way that

allows one to address them pro-actively rather than following a significant distance in their wake.

Software does not act completely blindly, at least not intentionally so, even if it is obscured as a 'Black Box'. It is invested with the values, philosophies and goals of both its original creators and those, including other software agents, who enrol and deploy it within their own plans and activities. Human faculties of memory and cognition are routinely enhanced through digital agents. For example, with Google to hand there are many facts and details one need no longer memorise. However, such digital agents, software and networks often obfuscate as much as they reveal when they function and exist as 'Black Boxes'. It is important to delve into how and why this opacity occurs in order to understand how those very same tools can be used to clarify, resolve and reveal themselves and other complex technosocial systems within which we are enmeshed in day to day living.

Examples abound of the complex interplay between humans, software and networks resulting in events overtaking any immediate human ability to grasp them and many unexpected consequences. For example, monitoring software used by FirstEnergy in Ohio led to a 'race condition'[1] that caused a mass power outage in the U.S. Northeast in 2003 (U.S.-Canada Power System Outage Task Force, 2004). Similarly, Haystack, a tool developed to assist Iranian dissidents communicate and penetrate government firewalls unintentionally left digital traces enabling the regime to track and locate users (Morozov, 2011, p. 207). In some cases, the interplay is so complex that reconstructions are still taking place years after the fact in an attempt to understand what happened and when, such as the 2010 'Flash Crash' in U.S. digital stock markets (see Appendix 1 for a detailed description).

---

[1] This can apply as much to electro-mechanical devices used for monitoring and regulation. In software terms, however, it usually refers to one or more algorithms (or 'threads') try to access the same variable at the same time resulting in incorrect and sometimes looping updates to the variable in question.

Indeed, Spencer Burns' in (McCallum, 2012) discussed how miniature Flash Crash events occur fairly commonly, such as the crash in United Airlines' stock on September 6th, 2008. Unlike the 2010 Flash Crash, this event followed a particularly straightforward and obvious narrative. Yet it was caused by a failure of a key 'Black Box' – the Google News web spider. It encountered an archived story regarding the United Airlines' 2002 bankruptcy. The article had no date attached to it, so the bot attached the current date by default. It was pushed out on the feeds as a current story, causing (human) traders to panic and drop United Airline's stock from $12 a share to $3 within a matter of minutes. We forget that such 'Black Boxes' are there, and subject to fault or change, at our peril. The Google spider responsible for disseminating false information was no less real for the fact that the data was incorrect. It *pushed back*, with tangible force nonetheless.

Software and networks generate highly tangible effects in human terms, even if their typical operation is obscured from view, and across multiple domains. My emphasis here includes the operation of networks as a core component in our understanding, and experience of, these systems. These networks contain, enfold and enrol both human and non-human (digital) actors. In this sense I refer to them as *networked assemblages*. Whilst it is interesting to study some of this software and the sociotechnical systems they are part of in isolation, I argue that the networked aspect is an important dimension to include, especially when exploring narratives. The highly networked nature of these interactions needs to be included at the heart of this inquiry. Using the term networked assemblages is useful not only for foregrounding the networked aspect but also for distinguishing to some extent these extremely fast, dynamic, assemblages from much slower and less dynamic (though no less interesting in their own right) human-object hybrids such as traffic patterns on the high street, or bureaucratic relationships in a business.

Networked assemblages are also ontogenetic – constantly in a state of becoming. This dynamism does not occur passively. Just as software and networks are able to do work in the world, *work* is required for software to execute or networks to transmit and receive. As such I regularly deploy a further term, *worknets*, borrowed from Latour (Latour, 2005, p. 132), although used in a differing (if compatible) context. This term helps to capture the inherent dynamism in the systems under consideration and maintain it at the forefront of attention in a way that the term 'network' lacks. Network is still used, however for clarity here it refers explicitly to specific physical (non-human) computer networks in order to demarcate them usefully from the more comprehensive and human inclusive sociotechnical worknets.

One of the main 'objects' of study here are machines, or rather the programs that run on them. Given that where an assemblage is concerned, humans and machines are not necessarily mutually exclusive in many cases the most interesting assemblages are distinct combinations of humans and machines. They are *hybrids*. This latter term is used in a particular sense by (Latour, 1993b) and explored further in the literature review in the following chapter. For now, though the important point is that 'collections' of humans and machines can be circumscribed. The concept is necessarily fluid. For example, in order for electronic trades to take place, humans have to interact with one another (and with '(ro)bot traders'), largely mediated by machines that carry and process communications across large, fast networks. Yet any specific combination of humans and machines in this sense is contingent and will change from day to day – indeed, from moment to moment as various deals are brokered with innumerable agents both human and non-human. Something mobile is at work. Assemblages may themselves move and transform, even if virtually - an entire trading desk can change focus from one market or segment to another for example. Entirely temporary assemblages can also be created, transformed, dissipate and reform across the

network. Describing all of the actors involved in a single trade for example, would be a single highly transitory assemblage. The mobility at work in relationships between components is both internal *and* external.

Capturing this kind of dynamic complexity is difficult. There are distinct epistemological and ontological difficulties underlying the descriptions of how communication and other activities, including research, are mediated with (and within) such networked and automated technologies online and also between both human and automated actors. This is particularly the case for attempts to describe such technologies at work *in situ*[2]. Dynamism, speed and lack of immediate human oversight or input contrived to hatch a powerful maelstrom of digital activity that had concrete real world effects. One of the most important features of computing related accidents such as the power outage in Ohio or the 2010 Flash Crash are that they are significantly driven by automated algorithms at work, responding to each other dynamically beyond immediate human purview. A more important issue however is whether they *could* have even been within the grasp of human oversight. The speed and complexity with which these particular algorithms interacted was, quite literally, beyond (immediate) human cognition. It is somewhat disarming therefore in this case to proffer a simple definition of algorithms in terms of decision trees even if is an accurate characterisation. It hides the complexity of both how the algorithm may behave with the extremely wide variety of inputs it may encounter from outside, including and especially those provided by other algorithms. As Steiner puts it, such rudimentary definitions give "little justice to the colossal webs they have become thanks to computers" (Steiner 2013, p. 128).

---

[2] 'In situ' in this case referring to describing how the technologies and users (human and non-human) are interacting at any one point. Traders customarily rely on a constant feed of market data and news aggregation. However, they do not necessarily have any kind of live awareness of changing relationships between trading systems, users, markets, strategies and so on. What is fed to them on via the tickers is 'reality' for them, irrespective of any fundamental underlying changes.

Whether the system in question is small or large[3] in relation to other systems has an effect on existing and possible relations between them. This is usefully illustrated by a popular term in the typical ICT developer's lexicon, "scalability", referring to the prospective capacity of a particular application or hardware configuration to be scaled up to deal with a much higher load of data, connections, users and so on. Typical conventions for describing human relationships with objects and one another (e.g. studying how humans may interact socially in buildings with different configurations and furniture), don't scale up well to domains that can bring us events like the Ohio power outage or the stock fall caused by a Google web bot. Conversely, common ICT practices for stress testing software and hardware[4] don't 'scale down' well to take account of unusual human behaviours. 'Zero day' exploits[5] of software for example are generally found by persistent hackers who think of novel ways to interact with software. Our conceptual furniture may need more enhancements and additions to help us usefully apprehend specifically, or even characterise generally, such assemblages[6].

Non-humans have, of course, always acted regardless of the presence of human. Indeed this was fully appreciated decades ago in computing by the early cyberneticists (Wiener, 1954; Ashby, 1957). Wiener explicitly named this process, where elements of human cognition and action were transferred to machines, "delegation". The primary difference now is that *synthetic* non-humans, built and deployed by human agency, are now capable of sophisticated levels of action with both humans and eachother (also a relative distinction of course,

---

[3] The relative size in question is also necessarily determined in an arbitrary way. If discussing (as we are here) a web of connections, where one segments a portion of that web off from the others is often arbitrary. For example, the computers I have at home (my 'home network') is a somewhat arbitrary selection of computers that are nevertheless all connected to the wider world network through the internet.

[4] A set of methods used to push the system beyond its original design parameters, usually with brute force methods (e.g. extremely high levels of data throughput), in order to identify weaknesses. Also related to 'Unit Testing', where the code is tested with particular sets of modules and sample inputs.

[5] These are vulnerabilities discovered in software that remain unpatched. In general usage this also refers to such vulnerabilities already being exploited by hackers who are aware of them before the software developer is, thus leaving open a window of opportunity. In some cases, these are vulnerabilities intentionally introduced by the manufacturers or developers (such as to ease eavesdropping activities by intelligence agencies).

[6] Particularly with regard to complex – and quite possibly dispersed - hybrid human-machine configurations.

however the degrees of freedom of a typical personal computer are a quantum leap away from that of a pulley system). This legion of synthetic non-humans, moreover, is now inextricably linked to the functioning of human collectives, especially in worknets. Far from subject and object being more clearly delineated, more clearly cut, and further distanced from one another than ever, we have on the contrary an ever deeper intimacy between the two, often in the form of a 'Black Box'.

Black Boxes

What are 'Black Boxes'? They are practices, processes and objects. According to Latour's definition something becomes a 'Black Box' when it

> …is made invisible by its own success. When a machine runs efficiently…one need only focus on its inputs and outputs and not on its internal complexity. Thus paradoxically, the more science and technology succeed, the more opaque and obscure they become (Latour, 1999, p. 304).

I claim that Black Boxes, following Latour's definition, are deployed in order to do two things, each of which is important to distinguish and where the terminology can be used as a verb or noun:

i) To effectively 'park' a complex series of relationships into a concept that can be used in a non-controversial way because it 'works' – for example, we customarily 'Black Box' (verb) our car engines into a singular well behaved concept of 'engine' and don't think about their internal complexity unless they go wrong in some way. This occurs in technical usage also. Much coding makes exactly this kind of simplification.

ii) To obscure a complex series of relationships that *is* controversial and poorly understood. Usually done in order to facilitate communication and to avoid being bogged down in a process of continual definition so that a wider context can be considered.

Evgeny Morozov critiques both uses. In particular, he argues that without sufficient awareness of the process occurring, the uncritical use of Black Boxes necessarily leads to excessively ahistorical accounts of current technologies (Morozov, 2011, 2013). Despite such concerns, the concept has proven a popular descriptive tool and technical practice across disciplines from the post war period onward. Cybernetics is probably the first identifiable domain to fully flesh out the concept and subsequently inspire later multi-disciplinary use. How the concepts are deployed to describe – for opening or 'whitening' Black Boxes (Glanville, 2002), or to dismiss networked assemblages (to 'close' the Black Box) is a core issue for the aims of this thesis in explicating how we can effectively apprehend complex networked assemblages. A reminder of their existence tends to occur in a dramatic fashion when they break or malfunction. This applies equally to relatively simple assemblages as it does the more complex. Plumbing may be taken for granted until it is blocked just as an entire trading network may be until it crashes. In this sense, the Black Boxes we interact with matters a lot, especially for customary day to day purposes. Those techniques that typically engage in 'whitening' are obviously preferable where it is practical to routinely deploy them. Even if they do not explicitly open the Black Box, reminding us regularly that Black Boxes are there that *could* be opened is an invaluable trait. This is another reason why the term *worknet* is useful in preference to network in such discussions.

Originating as an electrical engineering term (Cauer, Mathis and Pauli, 2000), Black Box was very shortly picked up in cybernetics with little change in its overt meaning; only altering its range of possible referents. It originally applied simply to a device or system solely in terms of its inputs and outputs. For engineers it was an electro-mechanical affair. For the early

cyberneticists it was significantly more inclusive, encompassing both the social and natural worlds alongside discrete physical devices (Wiener 1954; Ashby 1957; Mumford 1967). The term itself however was not annexed to cybernetics, even if the underlying concepts were there in Wiener's work, until (Ashby, 1957). Ashby characterised the 'Black Box' initially as a 'problem' that then became a theory applied in domains wider than engineering. This is in contrast to more recent, ubiquitous, usage such as in (Latour, 1999, 2005), where it is treated as a stand-alone, if highly dynamic and autopoietic[7], concept. Ashby argues that:

> In our daily lives we are confronted at every turn with systems whose internal mechanisms are not fully open to inspection, and which must be treated by the methods appropriate to the Black Box (Ashby, 1957, p. 86).

This characterisation is notable because whilst faster-moving, more dynamic technologies were already appearing in the 1950s, including continuing refinements of the earliest electro-mechanical computers, Ashby's approach was closely aligned with that of the electrical engineers the term was initially borrowed from. For Ashby, it was possible to interrogate Black Boxes through straightforward iterative means, asking questions over and over such as, "What properties of the Box's contents are discoverable and what are fundamentally not discoverable?" (Ashby, 1957, p. 87). There was little anticipation in his writing that the Black Boxes in question may develop to such a point as to perpetually escape the grasp of such iterative breakdowns, which is the primary concern in this thesis.

A whole new order of sophistication was added in the later wave of cybernetics. So called "Second order" cybernetics (or 'cybernetics of cybernetics') subsequently led to the appreciation of continuous, meta levels of Black Boxing, like Matrioshka dolls, with cybernetics reflexively applied to itself and the focus shifted from the observed to the

---

[7] Meaning a system capable of reproducing and maintaining itself.

observing system (Glanville, 2002). The claims made initially in cybernetics, even with this new development on the basis of the Black Box were relatively modest. Claiming knowledge was only possible through a (possibly arbitrary) circumscription of opacity, where only inputs and outputs are considered. 'Second order' cybernetics turned cybernetics in upon itself. Knowledge of, and through, systems was intertwined with their construction and maintenance.

The notion of 'circularity' in second order cybernetics, (Glanville, 2002) being inclusive of both the ongoing relationship of observer-observed or subject-object and the continuous opening (or closing) of successive Black Boxes, prefigures ever more sophisticated approaches to Black Boxing such as Latour's 'Circulating Reference' (Latour, 1999) – both in how to describe it and in how to continue Black Boxing ever faster and more complicated systems. As Glanville puts it

> second order Cybernetics, understood as proposing an epistemology and (through autopoietic systems) an ontogenesis, is seen as connected to the philosophical position of Constructivism (Glanville, 2002, p. 1).

And indeed, social constructivism was the next major iteration in this process, contending with the latest advanced technology Black Boxes "of Bakelite, missile guidance systems, electric vehicles, expert systems in computer science, networks of electrical power generation" (Winner, 1993, p. 366) and so on.

Black Boxing is visible in less dramatic (and therefore, less amenable to notice) examples too. Delegated decision making systems also have direct impacts. (Wallach and Allen, 2009) for example, consider the 'MedEthEx' system built by Susan and Michael Anderson[8]. It is an ethical decision making system for medical cases. It infers rules on the basis of casuistic

---

[8] See the Anderson's site dedicated to the system here http://webcampus.drexelmed.edu/MedEthEx/index.html - last accessed 04/10/2014

reasoning, iterating over decisions made by experts in the past on identical or similar medical cases to whatever the system is currently presented with. Because of fears regarding the potentially fallacious casuistry at work in such a system it is constrained by a series of cascading principles or 'duties' (cascading in the sense that some duties could override others), such as 'do no harm'. Decisions made by such systems have very real consequences for any patients treated on the basis of their recommendations. Quite aside from the difficulty of predicting consequences especially those that may violate a principle of 'do no harm' in a medical context, Wallach and Allen wonder about how such systems may "erode the autonomy and responsibility of primary caregivers" (Wallach and Allen 2010, p. 128). Ethical decision making becomes Black Boxed into the MedEthEx, or similar, system. Indeed, that this is a perpetual concern in robot and computing ethics is something they indirectly articulate without using the term Black Box or a synonym

> Designers of autonomous systems who choose to ignore the broader consequences of decisions made by their machines are implicitly embedding a particular set of values into these machines (Wallach and Allen, 2010, p. 31).

That is to say, where design choices are concerned even seemingly doing nothing, is effectively doing something: If we don't pay active attention to our continual Black Boxing of the complex and controversial into the simple and anodyne we risk misunderstanding or misusing the very technologies we build and deploy, not to mention incorrectly trace and ascribe responsibility for action.

The term Black Box since the Cyberneticists coined it has evolved well beyond an opaque system with inputs and outputs. Instead it is better understood, especially in Latour's usage, as a range of diffused effects distributed across a networked structure of human and non-human (for the current purposes, 'machine') interactions. This is even the case where the

clear distinction between the two becomes ever increasingly difficult to maintain, requiring a more fluid approach to characterise them. Latour straddles this middle ground area with his concept of 'hybrids' (Latour, 1993b), whilst Floridi uses 'inforgs' (Floridi, 2007). Both are grounded in the concepts and concerns first raised by Cybernetics.

Latour's distincton between 'technique/technology' and 'technical' is also useful for apprehending our interactions with Black Boxes. The former, he asserts, presents difficulties when used, as it usually is, as a noun. It often *masks* the activity and event of humans combining with various assemblies of both humans and non-humans. As a verb however, in the form of 'technical', it provides a useful adjective function highlighting our practice of briefly acknowledging and opening Black Boxes. "Technical,"

> applies, first of all, to a subprogram, or a series of nested subprograms, like the ones discussed earlier. When we say 'this is a technical point,' it means that we have to *deviate* for a moment from the main task and that we will eventually *resume* our normal course of action, which is the only focus worth our attention. A Black Box opens momentarily, and will soon be closed again, becoming completely invisible in the main sequence of action (Latour, 1999, p. 91).

The term frequently occurs when an obstacle or snag is hit in the 'main program' – i.e. the objective of any particular actor or 'actor-network' in Latour's terms. Indeed, it may even challenge the primary goal itself and perhaps lead to an entirely new course of action. If a virtual object created in Object Oriented Programming does not behave as expected, a programmer must 'open' the Black Box and inspect the object class's methods, arguments, constructors and so forth (see Chapter 6). If the object does not carry out the task required another must be found, or programmed from scratch or the original task revised completely.

Opening Black Boxes: networked assemblages and worknets

This early work carried out by the cyberneticists laid important ground for understanding the implications of automatic systems. Subsequent contributions of Social Constructivists (Huges, Pinch and Bijker 1989; Law and Bijker 1992; Bijker 1997), and then Actor-Network theorists (Latour and Woolgar 1986; Latour 1993) (re)introduced the human and social elements of technology development and deployment. One of Latour's most useful innovations in this respect was recasting networks as dynamic webs of black boxes that contained both human and non-human elements (see the literature review in Chapter 2 for more details).

When Latour first developed his notion of Actor-Networks in the 1980s, computer networks were still in a fledging state and mostly limited outside of research and defence domains as modest and highly localised business networks containing a handful of machines. Similarly, digital automation was the preserve of a very few practitioners. These early collisions between humans, digital networks and digital automation were highly significant however for shaping where we are now, with incredibly fast networks and automated agents ubiquitous in almost every aspect of our lives.

One of the most curious examples of these key 'collision' events, and a direct and enduring prelude to the 2010 Flash Crash referred to above, is detailed by (Steiner, 2012) - the case of Thomas Peterffy. He was the pioneer of automating the trade floor, first with algorithms that were followed by humans to beat the market as early as 1969 and then later fully fledged 'bots' that carried out trades autonomously at lightning fast speeds. Peterffy's story is interesting here because in 1987 his 'bots' were effectively censured by a Nasdaq representative, who gave an edict regarding their behaviour that still affects many 'bots' and their creators today – and web spiders in particular when faced with the 'robots.txt' directive

file (see Chapter 4). Referring to the (Nasdaq connected) terminal used to carry out Peterffy's automated trades, the official stated that it "...needs to be disconnected from this IBM and all orders need to go through the keyboard, typed one by one – just like the rest of our customers" (Steiner, 2012, Loc. 251). The intent was obviously that Peterffy replaced the machine with a human operator ("just like the rest of our customers"); however, Peterffy did not see it that way. He stuck to the letter of the edict by violating its spirit – two weeks later when the Nasdaq official returned, the terminal was disconnected as requested. However instead of a human sitting at the desk labouriously typing out trade after trade, Peterffy and his team had constructed a mechanical robot that read the output from the terminal screen via a camera, passed the data through an OCR application and then finally onto Peterffy's original algorithms. The master algorithm then sent orders by way of a series of pistons that hammered away at the keyboard.

According to Steiner's account, following his second visit, the Nasdaq official did not forbid the alternative contraption; perhaps anticipating that his words would be twisted similarly again. However, his objection is still commonplace to this day – there are some tasks that either should only be carried out by a human or if carried out by a bot, should be purposefully handicapped. How this applies in practice is an open controversy in almost every domain where 'bot' activity is visible enough to have a discernable impact. And whilst it is often difficult to justify restricting actions across the board to humans where 'bots' could save significant time and energy (for example, the 'bots' forbidden from using search facilities in some of the robots.txt files discussed in Chapter 4), requiring some kind of hobbling of capabilities is essential in others. Such 'hobbling' is at the core of recommendations resulting from the investigations into, and simulations of, the 2010 Flash Crash (CFTC-SEC 2010a, 2010b), in many cases requiring the slow down of market trading to prevent High Frequency

Trading bots from accelerating a market movement too quickly for human traders to adapt and respond to.

What emerges from such examples goes somewhat beyond just restating how humans and digital technologies (networks, 'bots') are interrelated and intertwined. They also highlight how some solutions to apprehending and describing networked assemblages and worknets can be usefully assisted by using the very computational and networking capabilities that are the subject of concern – something I touch on further in Chapter 3.

I propose here a term to encompass these kinds of 'mixed' solutions in terms of the work they do: 'elastic narratives'. In Chapter 3 I go into depth regarding two such theoretical approaches that I argue have significant and complementary strengths, both accounting for the dynamism inherent in today's networked assemblages: Latour's concept of 'Circulating Reference' (Latour, 1999) and Thagard's version of 'Computational Coherence' (Thagard, 2000). Why *elastic* narratives, though, as a generic container for such approaches? The answer to that is a compelling argument drawn from recent work on the impact and implications of 'Big Data'.

'Big Data' and Elastic Narratives

Big Data is a relatively recent entry to the technology lexicon. It is usually defined as data where the relative size of the dataset itself becomes an issue due to memory, processing, time or even conceptual constraints. Numerous refined definitions include other properties such as the 'velocity' (speed of arrival) and 'variety' of data. Anything from 'three Vs' to 'seven Vs' are proffered as defining characteristics – usually picked from an alliterated choice of volume, velocity, variety, veracity, variability, visualisation and value (van Rijmenam, 2013). What all of these definitions have in common is an appeal to extreme dynamism (often in the

form of complexity) and speed. The driving force behind this is the steady progression of computing technology, continually increasing the scale of storage, processing and communication speeds at ever cheaper costs[9].

Information overload has become almost omnipresent in the internet or information 'age'. And also something of a cliché. In many ways the practitioners of Big Data techniques aimed at solving Big Data problems are attempting to address this primary concern. For example, with early iterations of personal computers it may have been possible to know and understand all of their core processes and functions with the knowledge contained in only one or two books. Computing technology has advanced so significantly in size and complexity since then one would be lucky to own a computerised device whose range of software functions from top to bottom could be summarised in just two books.

Furthermore, the scales and speeds at which both the technologies, and the data, operate has made entirely new operations possible (such as, for example, Google's 'Map/Reduce' algorithm[10]). In short, both our relationships with these technologies have changed as have their relationships with one another. This is something that is possible to describe in general terms even if the added aspects of speed and dynamism are somewhat elusive in any initial effort at either description or prediction of behaviour. Theorising ways to apprehend and describe these aspects are tackled in Chapter 3 and elucidated further in the case study chapters that follow.

However, such attempts at creating accounts that capture dynamism and speed are only possible on the basis of being able to describe the underlying relationships thought to be at work initially. It is important to say 'initially' here as part of the problem is that these

---

[9] There is thus something crucially relative in any conception of 'Big Data', where 'Big' is only relative to the current capacities of computing technology in terms of storage and speed.

[10] This algorithm allows for the parallel processing of enormous amounts of data across multiple distributed machines. See for example http://research.google.com/archive/mapreduce.html

relationships can change dynamically and suddenly. In some ways we need to grasp a *range* of possible relationships rather than simply designating one as *the* working relationship. Hence *elastic* narrative.

Drilling down into the specifics of how one piece of code, or automation, may operate alone whilst valuable and informative, may make less sense in a wider context without mapping out its external relations with other agents. These relationships are already complex enough to occupy one's full attention so we need to consider how to do this first. As intimated above, using Black Boxes is often used as a means to deal with complexities by moving them into an abstraction category that is easy to discuss and refer to whilst hiding details. Both the practice and study of Big Data processes is replete with them. The Black Boxes used by humans and machines are important objects of study in their own right – in that they need to be 'opened' for exploration. A key to understanding imbroglios such as routine information overload in the ICT sphere is examining the opaque elements that are taken for granted at the centre of both. Assuming that networked events (of both human and machine genesis) or information overload are just brute facts is a result of unexamined Black Boxes. They do hidden work and make decisions on our behalf that may or may not be valid. There is a normative imperative at work here in urging that more attention is paid both to the roles that Black Boxes actively play and with regard to their inner workings. The literature review, outlines some cognate concepts that assist in effectively describe and actively handle this process[11], along with the theoretical perspectives in Chapter 3, providing a conceptual baseline for the remainder of the thesis. Practical applications of these concepts and methods are then explored in the subsequent case study chapters focusing on more empirical aspects, beginning with the consideration of bot behaviour in a web context. The latter focuses on a specific assemblage of internet users, browsers, web spiders, websites and web masters invoked when the server

---

[11] For example, 'obligatory passage point', 'trials of strength' and so on, drawn from Software Studies, Actor Network Theory and cognate interdisciplinary areas.

side 'robots.txt' file is accessed. This exemplar of bot behaviour is a smaller scale and manageable example of the overall case. Big Data practitioners by contrast consider much higher stakes, exhibiting their own peculiar characteristics and exhibiting approaches that are enlightening in this context[12]. All of which illustrate a need for a methodical approach. One of these most striking aspects of Big Data issues is how the arbitrariness referenced above in circumscribing assemblages is dealt with. The sheer sizes, speeds and dynamics involved mean assemblages are created, modified, destroyed and moved at high speeds and continually varying scales, frustrating attempts to cast them in stone and as such sharp definition is often eschewed in favour of more practical methods that do not seek to prematurely fix what is being observed, calculated or constructed.

It is made clear in numerous discussions about Big Data that being able to comprehend these relationships goes beyond just quantification and general concern with relative quantities. A number of scholars, commentators and practitioners concerned with Big Data (Boyd and Crawford 2011; Dumbill 2012; Lanier 2013; Mayer-Schonberger & Cukier 2013) have argued that the already common conception of Big Data is something of a misnomer: The idea that it was primarily the sheer quantity of information available growing to a point that computer scientists needed to develop new tools and technologies to store and process it is, for them, only one aspect of big data, not the defining crux. And this is the case, no matter which of the three, five or seven 'Vs' one selects as fundamental characteristics.

Big Data also means doing new things with the data we already have. New value is extracted from data sets that have gathered dust on the shelf, many of which could be quite modest. Rigid hierarchies are sometimes dispensed with. The design of database systems is being

---

[12] Worthy of note here is comparison of assumptions made by, for example, (Mayer-Schonberger and Cukier, 2013) that are challenged by (Boyd and Crawford, 2011). The former represent what could be considered the popular assumptions of 'Big Data culture' (i.e. the *culture* of Big Data practitioners), whilst the latter represent a thoughtful undercutting of trends that the authors already perceived taking over in said culture, such as the argument presented by Mayer-Schonberger and Cukier that sampling is now a thing of the past in the light of the amount of data available in Big Data.

revolutionised when required, moving from the pre-specification required by SQL databases to a new family of 'NoSQL' (Not Only SQL) database technologies where schemas do not need to be pre-defined and categorisation is fluid, dynamic and constant (Macreary and Kelly, 2013), although standard structured SQL databases still dominate. In these more unstructured cases, narratives in the forms of schemas are not set in stone beforehand but are allowed to follow the data wherever it goes or wherever it came from. Skilled data analysts are, effectively, creating new value with old data that otherwise would have been (in many cases, was) discarded or not collected otherwise (Mayer-Schonberger and Cukier 2013; Dumbill 2012b). And in so doing they are generating, reconfiguring and patterning new relationships not only between themselves and the technologies but also internally within the data itself.

As a result of this, the application and understanding of databases easily becomes a Gordian knot of entangled ontologies and epistemologies. Every reconfiguration of a database potentially brings with it a novel ontology and epistemology even if the data itself remains the same, though of course it rarely is static[13]. That being said, NoSQL approaches only make explicit the phenomenon that also easily afflicts strictly structured and administered schema based databases. There are still significant dynamism issues in the latter cases that are often difficult to encapsulate. Referring to even a schema based database as a singular object is problematic for example, as the data is often in continual flux and thus the database 'object' changes from one moment to the next even though in practice it is treated as a static, unchanging entity, including in programming terms where it may be Black Boxed as an object to query and manipulate via inputs and outputs. Similarly, when databases reach an unwieldy size they can 'shard' into smaller databases to ease the computational and server burdens.

---

[13] See for example, an introduction to the concept of 'NoSQL' databases, 'Making sense of NoSQL' (Macreary and Kelly, 2013).

This brings the inherent mobility and arbitrariness in the notion of assemblage back into focus – both computational and philosophical ontologies are being *imposed* and not just *found* in such data sets. Latour's 'Circulating Reference', explored in Chapter 3, proposes a consistent way to identify and trace such ontologies and, moreover, to do so in a methodologically useful way. Fixed definitions are of little use for describing dynamic entities and assemblages that can, quite literally, change in the blink of an eye. Identifying and tracing relationships is an important place to start in the tasks of apprehending, describing and understanding networked assemblages and worknets, bearing in mind a need to continually revisit given the dynamism and speed at work. This notion of continuous movement within a certain range is part of the motivation for referring to *elastic* narratives in these cases.

The emerging culture of Big Data practitioners highlights an important philosophical shift that has occurred in how we, collectively, view data. Even with the advent of the dot-com boom (and bust) and the ever increasing ubiquity of internet connectivity through the first decade of the 21st century, the size of data sets has been modest and their general availability limited. At best, secondary use of data has up until recently in the last decade been made by some more insightful data analysts using it for time series analysis. Due to modest size and similarly modest goals in imagining how the data may be used or combined with other data sets these would tend to be linear trends of best fit and therefore of limited use. Circles, cycles, waves and even more complex patterns may in fact more accurately model the phenomena being studied. Looking for linear relationships was the least computationally expensive operation in the past – a legacy we still contend with now. Computing power has advanced significantly, yet for many disciplines and enterprises to remain intelligible there must be some kind of continuity of method and historical data series. Similarly, the standard approach to data analysis (and collection) usually required the implementation of a schema

first within which the data would be expected to fit. This is a second aspect that is addressed by the theoretical consideration in Chapter 3 of 'Computational Coherentism'.

Practical approaches for apprehending and processing Big Data welcome a fluid approach that is both highly creative and reflexive. A dramatic example of these kinds of creative, reflexive range of approaches is a turnaround in *large scale* data re-use - Google's flu-trends analysis[14]. Google reprocessed search queries that would previously have only ever been processed once and possibly discarded by their search engine competitors. It was discovered, almost accidentally[15], that owing to sheer volume, user search queries into Google could be used to track infection rates and distribution of flu across discrete regional areas. Moreover, it was able to do so before professionals in the field from the CDC[16] were able to verify such infection rates on the ground. Just as interesting as why and how Google's 'Flu Trends' succeeded is also the mystery of why it suddenly failed too. From 2009 onwards, U.S. trends measured by more traditional measures diverged sharply from Google's estimates. Whilst Google representatives maintained that it remained reliable for the remainder of the world it is a curiousity that its estimation of U.S. trends would suddenly go awry.

Previously data, especially large amounts of it, had untapped re-use value and could come alive with a new dynamism when combined with other data sets in novel ways. This recombining was the typical method of data reuse when it was reused (rather than just occasionally referenced) at all. Mayer-Schonberger and Cukier make a particularly interesting argument that Big Data necessarily entails at least a partial rejection of seeking causal explanations in our epistemological approach, exchanging them instead for correlation. In their words – "not knowing *why* but only *what*" (Mayer-Schonberger and Cukier, 2013, Loc. 178) . It further highlights the passive intermixing of various ontologies and

---

[14] See http://www.google.org/flutrends/ for the latest information on the project. Last accessed 20/11/2013
[15] Through the application of a series of statistical models.
[16] Centre for Disease Control.

epistemologies visible in both Big Data problems and solutions (such as through the use of NoSQL). The ontologies and epistemologies at work are not always spelled out clearly. Part of this change comes from the *kinds* of epistemological and ontological definitions (exemplified by Mayer-Schonberger and Cukier) underlying the solution of Big Data problems. Algorithms designed to handle enormous throughputs of real time data often only require a rough level of accuracy. And in the case of real time applications such as tracking (and predicting) Twitter trends, this is all they could do anyway. Eschewing precision gives way to a range of outcomes, options and predictions providing *elasticity* to what may or may not be 'known' at any one time.

One of the most interesting insights from this more flexible approach to data is the argument that constraints that are assumed to be real may in fact simply be a result of the scale at which one is operating. Mayer-Schonberger and Cukier make the case that precision goes hand in hand with smaller data-sets. Further, they argue, this is difficult or sometimes actually impossible to achieve with larger data sets (especially those with real time characteristics such as Twitter feeds) and may in fact also be undesirable. Google's flu-trends and translation services show that there is a floor to certain types of data analysis (in terms of a minimum volume) that needs to be reached before any coherent signals or patterns can be identified. Computed translation techniques requiring raw pattern matching only became successful when the level of raw data available to correlate moved from thousands of data points to hundreds of thousands. One of the most striking floors (or ceilings) to be reached in this context is the collection in some cases of the entire data set for the population or question in hand. Google, for example, has access to complete search profiles as some users only ever use Google's search engine and no other. Given that in such cases *all* of the data can be processed, there is no longer any rationale, Mayer-Schonberger's and Cukier's argument goes, for sampling if the population remains relatively constant.

Sampling, Mayer-Schonberger and Cukier argue, is itself an artefact of information *scarcity*. The relaxation on standards of precision follows from this also. They refer to the need for precision as a way of thinking derived from a "small data" environment. Exactitude is a function of curation (selection) – where it is possible to perfectly curate the data (such as how many pennies are in your bedside drawer); it is possible to count with precision. They note that whilst researchers may not give up a desire for exactitude completely, it necessarily takes something of a back seat because not every specimen, subject or entry can be examined (especially if observation or data collection is delegated to automated agents). Historically speaking they argue that our difficulties in the past related to interacting with data were unfortunate artificial constraints imposed upon us at the time not be 'reality' but by the technologies available at the time. This is, again, where some level of arbitrariness has to be expected and somehow allowed for, or modelled in any descriptive methodology. The line between exactly what is imposed and what isn't is going to be continually fluid, and it is something that is expressed particularly well by Software Studies scholars (Berry 2011; Fuller 2006) as well as Actor-Network theorists (Latour 1999; Woolgar 1991)

Despite this ambiguity, one direction of travel is quite distinct: Noting that the size and 'messiness' of the data now collected naturally moves one away from achievable exactitude and with it, precise hypotheses regarding causation. Mayer-Schonberger and Cukier point out that the trade off in exchange is that the very volume and heterogeneity (especially in many cases, of sources not just type of data) enables connections and patterns to be perceived that may not have been hypothesised prior to 'letting the data speak'. That is to say, yet another important shift is the realisation that sometimes the hypothesis may come *after* the data is collected, not before. Boyd and Crawford, (2011) criticise this approach, pointing out that

it is the kind of data that encourages the practice of apophenia: seeing patterns where none actually exist, simply because massive quantities of data can offer connections that radiate in all directions (Boyd and Crawford, 2011, p. 2).

Whether one in fact is a victim of apophenia or whether one is actually detecting real patterns that identify some kind of independent relationship between variables is likely an impossible distinction to make *a priori* without some kind of viable external adjudication. It also doesn't help if there are competing ontologies and epistemologies to be considered without some means to distinguish between them. These questions are considered later in Chapter 3 when discussing 'Circulating Reference' and 'Computational Coherence' and in the case studies in Chapters 4 and 5.

Furthermore, whether one is more optimistic or pessimistic regarding the possibility of detecting genuine patterns in Big Data, either position has important implications for the notion of expertise – at least on an individual level. A data scientist with a sufficiently large data set and the right tools may be able to glean insights beyond the ken of a subject specialist who relies more on their experience in the field. Mayer-Schonberger and Cukier claim that experts won't disappear, however they will have to contend with the distribution of epistemological tools in their domain that will force adjustments to their own analyses even if deployed by researchers who are only laypeople in their specialist field. Winners of many Kaggle competitions[17] for example often hail from very different disciplines to those for which their programmed data solutions glean insights. This has important connotations for institutions as we currently understand them, especially those whose primary *raison d'etre* is knowledge production, collection or communication, as Mayer-Schonberger and Cukier point out

---

[17] http://www.kaggle.com/competitions – Regular competitions used to solve data science challenges. Winners are encouraged to tell prospective data science employers about their success in competition as it is fast becoming a de facto standard for picking out talented data scientists.

Most of our institutions were established under the presumption that human decisions are based on information that is small, exact, and causal in nature. But the situation changes when the data is huge, can be processed quickly, and tolerates inexactitude. Moreover because of the data's vast size, decisions may often be made not by humans but by machines (Mayer-Schonberger and Cukier, 2013, Loc. 264).

Methodologically speaking, software studies and cognate disciplines already anticipate this kind of dynamism. However, in order to even reflect such dynamic activities may itself also require the help of machines and not just the personal hand written notebooks recommended by (Latour, 2005) to capture the intricacies of the network being studied. The overload of information quantity that Big Data techniques seek to address would obviously go beyond the scope of a handful of researcher's notes seeking to trace the relationships at work. There is much cognitive outsourcing now (Shirky, 2008), with machines acting independently not just as calculating centres or decision makers. We also regularly collectively delegate the task of observation to them. As Mayer-Schonberger and Cukier note above, this is now embedded at an institutional level. Therefore, I argue in this thesis that it is a key part of any description of such technologies in action to also identify these new relationships in a way that captures the speed and dynamism at work but is comprehensible to humans. There is little to be taken for granted in this respect. Furthermore, such issues are not just limited to the processing of colossal amounts of data. They also impinge directly on how we characterise our relationships with physical machines and decision making systems. The most appropriate approaches, I argue, have a *heuristic* character – they *tend towards accuracy*, but don't guarantee it.

Case Studies

In order to explore the theoretical work discussed in Chapter 3, in the second half of the thesis I work through several case studies. A brief introduction to each is provided in the chapter summary below. However, it behoves me to note two interesting developments that occurred during the process of collecting the data for the first two case studies (Chapters 4 and 5).

What makes them interesting is that they are broadly illustrative of the overall problem domain that this thesis seeks to contribute solutions to. It has already been noted above that we have difficulty in apprehending fast moving networked assemblages, especially due to the general limitations of human cognition. Changes in these assemblages can also be sudden and unexpected, as I found out directly. In the first case, whilst engaging in the research one of the data sets I was making use of fundamentally changed its structure and in the second, one of the automated tools I sent data to for processing neatly reversed its methodology.

I investigate and interrogate multiple 'robots.txt' (directive) files in Chapter 4. Whilst collecting them and deciding upon candidate sites, the individual files that originally existed for specific government departments were rolled up into a single 'robots.txt' file on a single government site. This was the 'structural change'. The second was the 'churnalism' tool based at 'churnalism.com', relied upon for the work in Chapter 5. Originally the tool functioned by comparing chunks of input from press releases to the entirety of the 'journalisted' database which contains the majority of UK online mainstream media output. Part way through my research, the tool's creator unexpectedly altered this process completely. The new version instead collected press releases from a selection of

organisations and then compared these to inputs taken from media clippings. It was more or less complete reversal. This meant that the originally intended research was only partly completed – resulting in a comprehensive amount of data gathering for two government departments (DEFRA and the Environment Agency), yet being left high and dry with a long list of other potential sources of data I had drawn up including other government departments and the top ten businesses listed on the FTSE stock index. Bots had been constructed and deployed successfully and a large amount of data was returned by the bots – sufficient to inform Chapter 5. However, it was only of limited use for studying a range of imbroglios in which individual bots may find themselves. The flashpoint in each case was the assemblage of bot automatically requesting and returning information, the configuration particulars of the server it interacted directly with and the possible attention, even interference, of a human webmaster responsible for maintaining the site and possibly the server it was hosted on. My treatment and discussion of the research was adapted accordingly to these two developments.

Following those caveats on the case studies, and before moving on, I will briefly explain the structure of the following chapters.

Chapter Review

Chapter 2: Literature review

The practice of Black Boxing features largely in this review. Its genealogy is considered given that it has a long pedigree that is used both as a conceptual shortcut (avoiding the deep and complex analysis required to accurately describe a sociotechnical state of affairs for example) and as a rhetorical tool. The thesis is placed in the context of an ongoing debate surrounding the status (and 'practice') of Black Boxing.

This is then further situated within more contemporary explorations of human and technology relations in Science and Technology Studies (STS), Software Studies, Actor Network Theory (ANT) and other cognate schools of thought. Particular concepts from Latour are situated within the ANT context and then drawn out as they form the basis of much of the discussion in the remainder of the thesis. Limitations of both contributions from ANT and Latour are considered.

The body of literature supporting 'Coherentism', the second major conceptual resource for this thesis, is then covered, including the key debates between 'Coherentism' and 'Foundationalism' in epistemology. Finally, once the key concepts of the thesis have been situated more fully within the literature a taxonomy of bots is provided to provide a clearer characterisation of some of the objects of study.

Chapter 3: Theorising the Computational and the Digital

The concepts of 'Circulating Reference' and 'Computational Coherentism' are key to the remainder of the thesis and considered in depth in this chapter. The two key aspects, initial description of assemblages and the activity underpinning them, is something of a circular relationship. 'Circulating Reference' is presented as one of the primary heuristic methods offered to address the difficulty of apprehending complex networked assemblages. The concept is adapted specifically for the task of creating the kind of 'elastic narratives' that could map out the networked assemblages and their attendant epistemologies and ontologies.

Given the inevitability of conflicting epistemologies and ontologies in such accounts, not to mention the raw speed involved in networked assemblages, a further approach is required that is able to sift between competing accounts and also do so in a timely fashion when applied

with the assistance of a computer. 'Computational Coherentism' is proposed for this task and in particular the work of Paul Thagard, is considered.

'Circulating Reference' assists with the ontology – the 'story' describing who, what and where in a dynamic, fast moving networked environment with less risk of spurious correlations and pattern identifications. However, it offers little help in the face of competing accounts and is soon out of its depth when dealing with computing speeds. Building on the previous chapters, 'Circulating Reference' is used as the prelude for coherentist accounts enabling a more informed choice between narratives. 'Circulating reference' allows for accounts that can accommodate changes over time. 'Computational Coherentism' is proposed as a further tool that can enable accounting for *fast* changes and provide a map to guide between multiple narratives.

Chapter 4: Case Study - Robots.txt

Collected 'robots.txt' directive files are used here to deploy some Latourian concepts, particularly Circulating Reference, in understanding a particularly common networked assemblage – the interfaced networks created between bots, humans, browsers, webmasters and webservers. This enables the adaptation of some Latourian concepts used throughout the thesis in real world terms and also highlights two of the main issues raised above – the need for proper attention to be paid to the epistemological and ontological accounts that underpin current accounts and the limitations in following such relatively static and post-hoc approaches without additional tools such as Circulating Reference that allows for 'elastic narratives' to be developed.

Chapter 5: Case Study - Churnalism

Talking about the interplay of bot, network and human tempts premature Black Boxing of elements that still remain controversial. This is especially the case as the sheer complexity and number of these interactions can lead quickly to narratives based on apophenia. The phenomenon of 'churnalism' (the widespread copying and pasting of press releases passed off as responsible journalism) is considered here as a particular problem in that it attempts to impose specific narratives in the chaotic digital world. How this impacts the digital ecosystem and how it can be responded to in digital terms highlights how both bots and their networked interactions with humans can only be understood and described in dynamic terms which require appropriately dynamic epistemologies and ontologies.

'Computational Coherence' is invoked to address both the problem of 'churnalism' used to open and situate the chapter as well as the wider problem of choosing between competing 'elastic narratives' when talking about networked assemblages.

Chapter 6: Case Study – Object Oriented Programming

This chapter ties in Object Oriented Programming with the overall approach set out by (Latour, 1999), further extending his concept of Circulating Reference to include the Latourian notion of 'Program' in order to adapt Latourian concepts even more fully for application to talking about digital objects in sociotechnical systems.

Chapter 7: Conclusion

We have great difficulty in anticipating the complex outcomes of networked assemblages. Their effects, their capacities and their blindspots. This final chapter ties together the contributions made in both the case studies and the general application of both 'Circulating Reference' and 'Computational Coherentism' towards addressing this, advocating these two as additional tools for effectively apprehending networked assemblages and making some proposals for further directions for similar research.

Chapter 2: Literature Review

The concept of the 'Black Box' is multifaceted. It covers a range of possible objects with real or virtual moving parts and is also a *process*. As a term itself, it has a provenance and history traced initially in the previous chapter through cybernetics. Many of the subsequent discussions and considerations of this thesis sit within the broad application of the Black Box and the activity of Black Boxing.

In addition to his particular take on the meaning of Black Box, several of Latour's other concepts carry out much of the heavy lifting for the subject matter here. To be fully intelligible, they are explained in terms of his main contribution to the social sciences, Actor-Network Theory (ANT). The latter developed in a much wider context – that of Science and Technology Studies (STS) and Software Studies, which are also briefly covered. Whilst the contributions of STS are not as essential to the direction of travel here as parts of ANT they nevertheless set the scene for scholarly inquiry into the social roles of the technical and the technical roles of the social that heavily influenced discourses on technology in the West during the last few decades of the twentieth century.

The Latourian concepts that are deployed here, such as 'Circulating Reference', 'Obligatory Passage Points', 'Trials of Strength' and so on, are tightly associated with him and sometimes only lightly associated with some of his collaborators such as Woolgar and Law. Nevertheless, they are not inseparable from his work in ANT and most, (Latour's direct outside philosophical influences aside), were insights gleaned from his and other ANT practitioners' careful study of complex sociotechnical networks, from Aeroplane manufacture (Law, 2002) to laboratory experiment regimes (Latour and Woolgar, 1986) to soil sampling in the Amazon basin (Latour, 1999). There are thus regular references to ANT throughout,

though it should be borne in mind that the work here draws not so much on ANT itself (beyond it being a key context) as it does on some of Latour's specific ideas.

To complement and update Latourian concepts for effective application in the fast moving digital environment(s), a substantial amount of time is spent later in the next chapter on considering Paul Thagard's flavour of 'Computational Coherence' theories. The most relevant literature for orientation in these areas is explicated in this chapter.

Finally, as the more empirically focused parts of this thesis looks at a much more simplified arena, that of web based bots and their numerous social and technical interactions with humans and other machines, a 'taxonomy of bots' is provided to aid the reader in making sense of the multiple references to bots and different types of 'bots' and 'bot' behaviour that are peppered throughout the thesis and are also the primary subject matter of the case study chapters.

The literature drawn upon covers several threads that are conjoined with the question of describing fast moving, dynamic sociotechnical systems – or to use the abbreviation proposed in the previous chapter, *networked assemblages*. Both kinds of phraseology are used interchangeably throughout according to the emphasis required. 'Networked assemblage' for example does not leave it obvious to the reader that a social element is very likely intimately involved, so 'sociotechnical system' may be more appropriate in some cases. Similarly, with the Latourian concepts in play sometimes the context means it is more appropriate to refer to them as Latourian (usually in a more philosophical sense) and at other times to associate them more closely with ANT or 'actor-networks' (usually in a more practical or 'deployed' sense). There are a few other similar disambiguations to be made that are introduced at appropriate points.

Black Boxes beyond cybernetics

Applications of the Black Box (as both *thing* and *process*) went beyond cybernetics and into the broad reach of social sciences and philosophy via STS (Science and Technology Studies). In STS accounts, the concept of 'machine' (where Black Box had originally been derived) itself is no longer regarded as a singular, isolated, object that opposes the human in some neutral void. Rather it can quite easily be taken to encompass both, taking the place of a neutral void and acting as an apparatus distributing possible relationships between distinct elements – whether they be human or other machines. This 'distribution' of capabilities or possibilities occurs through numerous proposed mechanisms such as Deleuze's 'lines of flight' (Deleuze, 1990), or Latour's 'Property Swapping' and 'Circulating Reference' (Latour, 1999).

'Machine' can thus be expanded to include discrete (if *arbitrarily* discrete) entities including forces, gadgets, networks, bodies and all manner of agents and their relationships. Deleuze and Guattari's use of *assemblage* encompasses this, which in turn is derived from Foucault's notions of *dispositif* or *apparatus* – a continuous, dynamic set of relations where technology and subjectivity (both very widely conceived) are not particularly clear cut (Gordon 1980; May 2005). Deleuze and Guattari do not provide an explicit and consistent definition of *assemblage* though they do often provide examples. In true reflexive terms for example, they refer to a book in *A Thousand Plateaus* as an example of an assemblage (Deleuze and Guattari, 1987, p. 4).

Similarly, notions of a Black Box make their way into the analytic philosophy literature – a mainstay of Floridi's Philosophy of Information for example, is 'Levels of Abstraction' (Floridi, 2006, 2010). It is a notion that bears great similarity to the various articulations of 'Black Box' above that emphasises a continually unsettled nature in the sense of a potentially

never ending chain of Matrioshka dolls to be found one within the other. This is in contrast to the more 'continental' approaches of Deleuze, Bijker and Latour that emphasise, (via notions such as assemblage, apparatus and so on), an ongoing active process. The definitions are not incompatible, instead choosing to focus on what is considered to be either the most important or most interesting distinguishing aspect.

STS is considered here primarily via its initial technological turn – drawing technology into accounts that were, in social constructionist terms dominated by humans interacting with technological Black Boxes that were always subsumed by social construction. STS subsequently drilled down into the latter in earnest (Callon 1986; Woolgar 1991; Latour 1999; Law 2002, 2004; Sismondo 2003).

A subset of STS, 'infrastructure studies' provides a useful bridge between the broader STS accounts and the comparably more focused digital material that is the focus of this thesis (Bowker et al. 2010; Star and Bowker 2002). In looking at organisational and technological infrastructure the latter provides some useful insight. A key part of which is highlighting how many computerised systems are in fact products of pre-existing paper based procedures.

The most substantial scholarly treatment of Black Boxing that attempts to tread a middle ground between social and technological determinism is Actor-Network Theory (ANT). ANT is used here as the largest conceptual base within STS from which the thesis draws in two respects:

i) As a methodology (Latour and Woolgar 1986; Latour 1988, 2005), for describing complex human and non-human networks, known as assemblages (Latour 1993b; Law 2004).

ii) As a comprehensive philosophical account, drawing on Latour and Whitehead's

contributions (Latour 1988, 1993, 1999, 2010; Whitehead 1947, 1960; Harman 2009, 2011), that enables description of 'networked assemblages' in ontological and epistemological terms. Whitehead provides the inspiration for much of Latour's underlying ontology, particularly in the form of his concept of 'Proposition', which describes the event or encounter of two entities (Whitehead, 1960).

Software Studies is a relatively recent inheritor of STS and ANT that reorients explicit focus to the digital (Fuller 2003, 2005, 2006; Berry 2011). Networked and automated technologies being both dynamic and extremely fast and tend to be amenable to the kind of layered, 'Black Boxes all the way down' analysis (e.g. Fuller, 2005) favoured by Software Studies scholars. However, as other parts of STS have refocused and adapted to the increasing ubiquity of digital technologies, the line between Software Studies, ANT and STS becomes increasingly fuzzy as both subject matter and approach tend to converge. Previously diverse subjects have become, at least in some individual treatments, indistinguishable from Software Studies. More recent STS and ANT accounts for example, refocus on communications (Paccagnella 1997; Boczkowski 1999) and digital technologies (Venturini 2010; Venturini and Latour 2010; Berry 2011; Boczkowski et al 2008; Bell 2010; Weiss and Domingo 2010; Bucher 2012).

Whilst it is arguable where a putative Software Studies discipline may have begun (some may, for example suggest (Castells, 2000) and his notion of the network society), the term is first used to refer to an emerging field by (Manovich, 2000). Software Studies is defined by (Dodge and Kitchin, 2011) as focusing

> on the etiology of code and how code makes digital technologies what they are and
> shapes what they do. It seeks to open the black box of processors and arcane
> algorithms to understand how software — its lines and routines of code — does work

in the world by instructing various technologies how to act. (Dodge and Kitchin, 2011, p. 13).

This 'etiology of code' is explored in multifarious ways in the literature. For example, (Galloway, 2004) looks at the role of code as a disciplining agent (for better or for worse) in the form of protocols. In a similar vein, (Lessig, 1999, 2006) asks whether code can be regarded as having a lawmaking capacity. Meanwhile, (Montfort et al, 2013) investigate the cultural impact and significance of a single line of code.

There are numerous allied developments elsewhere that help to inform this thesis, that also follow the theme of identifying and opening 'Black Boxes'. Scholars such as (Welchman 1997; Delanda 2011) drawing on Deleuze, provide examples of 'machinic' thinking and concepts arguing that 'machines' can be conceived, *pace* ANT, as containing both human and nonhuman, virtual and physical components.

Similarly, the informational and computational turns in philosophy are important cognates, if not necessarily allies, to the endeavours of STS for example (Allo, 2010) and (Bynum and Moor, 1998) respectively. They both refocus attention on the 'allocentric', meaning going beyond just human concerns and taking objects, in this case informational and computational objects, seriously. One of the most significant developments in this respect within the broader philosophy of technology and computing is the work of Floridi and followers, whose work has built upon and extended cybernetic accounts, (Floridi 1999, 2010, 2004). The latter also dominates the ethical discourse on computer and robot ethics with a handful of accounts (Johnson 2004; Wallach and Allen 2010; Sharkey 2010) running in parallel to the dominant Floridi and anti-Floridi discourses.

There is also a substantial body of literature originating in the study of technology from the perspective of law such as (Benkler 2006; Solove 2008; Johnson, Regan and Wayland 2010;

Lessig 2006). The developing sphere of technology law is both reactive and proactive with regard to the evolving uses of advanced digital and communication technologies and as such is highly informative. This applies as much to testing particular philosophical concepts 'in the field' as it does for highlighting edge cases that may force reconsideration.

These act as a springboard for better known debates in popular recent accounts of the business and technology literature. (Steiner 2013; Friedman 2007; Rheingold 1993, 2003; Zittrain 2009; Shirky 2008; Tapscott and Williams 2008; Kirkpatrick 2011; Singer 2011; Lanier 2011, 2013). Many of these accounts not only provide interesting narratives individually but also an object of research in and of themselves to the extent that they Black Box or unpack the technologies they consider.

More current focus on specific technologies and networked assemblages are found in considerations of Big Data, (Michel et al. 2011; Onnela et al. 2007) and similar research that anticipates potential crises in certain subject areas such as sociology that do not properly adapt to these technologies or their application (Savage et al 2007; Mayer-Schonberger and Cukier 2013). The recent work by (Venturini 2009, 2010; Venturini and Latour 2010) is also noteworthy as it begins the bridging of digital technologies and Actor-Network Theory, albeit from a different direction to myself. Venturini and Latour explore in these works how to use digital technologies as an effective means of *expressing* Actor-Network theory.

Actor-Network Theory and Latour

Latour highlights in his introductory text to ANT (Latour 2005, p. 63), that he encountered significant difficulty in communicating its approach. This is because the terms 'Actor' and 'Network' have acquired popular meanings since the ANT's inception in the early 1980s prior to the rapid development of the web and computer networks more generally (which at this time were still the preserve of academic and research institutions). By 'network' Latour

is at pains to point out that he means something that *could* resemble what we usually now associated with a 'network' shape, however this is not fundamental to its meaning. He proposes that in cases where both his, and the common parlance is to be used, 'worknet' should be used to distinguish the 'network' of 'Actor Network Theory' from the colloquial meaning of 'network' associated with computer and internet networks. The 'work' in 'worknet' comes from the notion that 'work' must be done for actions and changes to take place within the 'net'. One of the primary aims of ANT is to remove vague (and often damaging in both descriptive and explanatory terms) assumptions of 'action at a distance' – some kind of 'work' *must* be observed for something to count as a network/worknet, even if it is just (as may be the case in a computer network) an occasional heartbeat (ping) sent across the network to monitor ongoing connectivity.

The worknets of ANT are always in flux and, it is asserted, only through such changes can be observed at all as they leave 'traces'. Any and all objects can be considered 'agents' and part of a 'worknet' in ANT as well as humans even though Latour acknowledges that the former do not necessarily have any intentionality as such (Latour, 2005, p. 47–58) they do however have *effects*. In addition, they all form associations with other actors to form worknets which in turn link to other worknets also consisting of both humans and non-humans. It is important to note that no methodologically significant distinction is made between these two types of agent. Instead ANT theorists identify two types of observable agents – *mediators* and *intermediaries* (Latour, 1993b, p. 79–82). The former effect some change on a signal or behaviour whilst the latter simply repeat or transmit it.

An additional important feature linked to both the ideas of mediation and translation is that of *obligatory passage point*. Latour introduces this in *The Pasteurization of France* (Latour, 1993a, p. 43–44), using it to describe the passage that Hygienists in 19th century France were forced to take by the discovery of microbes. Inspired by military terms, he defines obligatory

passage points where "They [the Hygienists] have only to concentrate their forces at those points for their weakness to turn into strength."

An important corollary to the obligatory passage point is Latour's version of 'Institutionalisation', where Latourian 'Institutions' are themselves obligatory passage points. 'Institutionalisation' is the process through which the 'stabilising' of 'substance' occurs. It means affording the kind of historicity to non-humans that we are used to affording to human (social) events. As Latour quips, "there is a social history of things and a 'thingy' history of humans" (Latour, 1999, p. 18). Our idealised conception of scientific advances runs contrary to this idea. It instead supposes that something, once discovered, 'was there all along'. This is a key point at which Latour's realism collides with his constructivism. To accept it not only means returning to the age-old philosophical divides such as subject/object, science/nature, epistemology/ontology. It also means, for Latour, adopting a rigorous account of how one scientific theory surpasses another (Latour, 1993a, 1993b, 1999).

The properties of physical objects, such that they can be determined, appear in the context of tests, not as properties in and of themselves in isolation. Realism drives Latour's ontological conception because physical objects are able to "push back". If one thinks of human and non-human actors as equivalent, it is easier to see what Latour means when he describes physical objects of study as having "interests". Researchers in the 'hard' sciences attempt to isolate and manipulate the 'interests' of these objects, however they resist that manipulation and push back against the worknet trying to manipulate them (of which the scientist is part) (Latour 1993; Latour and Woolgar 1986).

For Latour, when sociology turned its gaze to 'hard' science, it encountered participants who would and could push back (in much similar fashion to their physical objects of study) in a way that participants in social science studies would or could not. Or were prevented from

50

doing so (usually by a sociological interpretation that 'explained away' objections). For ANT, the 'hard' sciences have become powerful because their translations tend to be rigid (unlike the social sciences) and alternative translations are easily detected and/or amended where necessary. This is further reinforced by the process of 'Institutionalisation', where work is carried out within the institutional worknets to maintain these rigid 'translations' (Latour, 1999). Latour's specific conceptions of 'Institution' and 'Institutionalisation' are used as a means to explain how, in a Latourian conception, sociotechnical objects or worknets can have any kind of inertia.

Latour's non-deterministic anthropological approach to epistemology is somewhat anticipated by (Winner, 1980, 1988) and heavily influenced by (Whitehead, 1947, 1960). Winner went beyond the exposition of just making and using technologies to include the dimensions of, for example, politics generating artifacts that "embody specific forms of power and authority". Latour surpasses Winner's conception though in arguing for genuinely fluid boundaries. Similarly, Latour's notions of *Articulation* and *Proposition* are borrowed heavily from (Whitehead, 1947, 1960), complete with the problems of Whitehead's occasionalism (Harman, 2009). 'Articulation' is Latour's technical synonym for description. The key difference is that any particular object, agent or state of affairs subject to description can have multiple articulations – and the greater the number of these articulations, the more 'real' it is. 'Proposition' carries out a conjoined role with Articulation along these lines. As Articulations replace descriptions in Latour's ontology and epistemology, Propositions replace 'statements' *about* affairs, or the typical use for the term 'proposition' in philosophy. To distinguish these two meanings, I use a capital for Latour and Whitehead's meaning and lower case for the conventional meaning when it is propitious to use it. A typical example of a proposition in the conventional sense would have the form 'x and y did x'. A Proposition on

the other hand, refers to the *meeting* of x and y, or rather the *event* of x and y encountering one another (possibly via a third party that *could* be x).

Another important concept is *trials of strength* occurring between, for example, parasites (Latour 1993a; Serres 1982) or in the digital era, computer programmes (Berry, 2011). Every agent has at least one direction of travel, or goal. However, when it comes into contact with other agents, they deflect one another's trajectory and often engage in a realignment – a trial of strength. Whichever agent is able to mostly maintain its original direction is the 'winner', though as Latour repeatedly points out (Latour, 1993a, 1993b, 1999, 2005), no such encounters (Propositions) occur without both parties and their "program of action" being changed, even if only slightly – and thus being *mediated* or *translated*.

The meaning of 'social' is often misunderstood in Latour. He makes his meaning particularly clear in *Reassembling the Social* (Latour, 2005): 'the social' and indeed, 'sociology' should, he argues, return to its roots in Gabriel Tarde's conception of the social sciences – that it is about the study of *associations*. A frequent critique deployed by Latour against the current social sciences is asserting that they adopt some kind of 'conspiracy theory' with a supervening and mysterious social 'stuff' acting as the prime mover (Latour, 1999, 2005), much like on the other side of the subject/object divide, the inertia assumed in 'objective' reality that would preclude actor-networks such as the institutionalisation of Pasteur's discoveries from having to act at all.

Latour deploys A.N. Whitehead's notion of *Propositions* (Whitehead, 1960, p. 280) to explain why this is the case. These Propositions are contrasted with *statements* that usually seek to establish a connection either through correspondence or resemblance between words and states of affairs in the world. There is not a single facet with which an object, Black Box, actor or actor-network can be summed, and grasped with a word, as many epistemologies

would have it. Instead, it is the *differences* that distinguish Propositions; differences that no one necessarily knows in advance including their number, relative sizes and disposition. It is because Propositions are *occasions* (or 'events') where two or more entities come into contact, they give the opportunity for such differences to become visible in some way.

Further, the way relationships are established between Propositions is through *Articulation*. Where Propositions (occasions) happen, objects meet and articulate one another, both humans and non-humans articulate eachother in multifarious ways as this action is limited "not only to words but also to gestures, papers, settings, instruments, sites, trials" (Latour, 1999, p. 142). Propositions are both *surprising* and *events* in the histories of other entities. And the more means of articulation available and deployed, the better. Non-humans are not mute – not in interaction with humans or with other non-humans. Therefore, "we have to abandon the division between a speaking human and a mute world" (Latour, 1999, p. 140). The entities under examination exist because they are articulated between so many others, including in many artificial settings such as laboratories. To put it another way:

> We speak *because* the propositions of the world are themselves articulated, not the other way around. More exactly, *we are allowed to speak interestingly by what we allow to speak interestingly.* (Latour, 1999, p. 144, original emphais)

When we go through the process of 'stabilising an entity' (i.e. effectively collectively agreeing to what it is), through Circulating Reference, Propositions and Articulation, we move from attributes to a substance. Something with limits, persistence. 'Substance' here designates the relative *stability*, where a multiplicity of actors are gathered into a stable and coherent whole. This substance is not permanent, but reliant upon this relative conjunction of encounters, networks, relationships and the work done to stabilise it.

Latour's 'Circulating Reference'

The clearest exposition of Latour's metaphysical and ontological position is set out in his book *Irreductions*. It is a separate piece tacked on almost as an afterthought to *The Pasteurization of France* (Latour, 1993a). I do not seek here to defend Latour's specific metaphysics, though along with sympathetic critics such as Harman, I do believe it is a compelling and defensible account in spite of its problems. Rather, it is some of Latour's insights that are derived from this manner of thinking that I wish to deploy in order to tackle some of the pressing questions raised in this thesis. They require some familiarity with his particular position in metaphysics however – or at least the position he once held. Whilst all of his works contain significant amounts of philosophical depth, the consistent and comprehensive bulk of his writing concerns broad scientific, social scientific and empirically focused observations and inquiries. As such it is difficult to be sure how much of his original stand as set out in *Irreductions* is still held by him. However, as Harman argues, "If Latour eventually abandons some of the claims in this treatise, we should first adopt them in order to share in their later abandonment" (Harman, 2009, p. 12).

Latour details in this book an epiphany he experienced that led to his primary insight that drives the remainder of *Irreductions*:

> Nothing can be reduced to anything else, nothing can be deduced from anything else, everything may be allied to everything else (Latour, 1993a, p. 163).

Referring in the same paragraph to himself and other 'objects' (of any variety one cares to imagine), he says, "we mutually defined ourselves". 'Objects' (or 'actors' or 'actants') can no longer be assumed from this point on in Latour's personal journey. They require a deep empirical focus, following them wherever they go. They have no fundamental or hidden

54

essence and are defined entirely by their relationships with other objects. Moreover, we cannot even perceive them beyond these relationships, or via other objects. This 'following', for Latour, becomes the seat for a serious empirical undertaking in all of his other work – there is little of note, on my reading of him, to be discussed beyond such empirical pursuit of objects, their relationships and their trajectories. This certainly fits the character of easily dissipated or interrupted relationships at work in digital contexts.

Generally speaking in philosophy, language and nature are regarded as entirely separate domains with epistemologically risky attempts such as theories of correspondence, representation – even epiphenomenon being deployed to bridge the purported gap. Latour proposes that *Circulating Reference* more accurately describes how words and world relate (Latour, 1999, p. 24).

The hard (or digital) copies of our inscripted representations in the form of the written word are what Latour calls 'immutable mobiles' (Latour, 2005, p. 226). They can be and are circulated and manipulated beyond the contexts they are derived from. Nature is essentially transformed, representationally speaking, into marks on a paper or bytes in a computer. These inscriptions are manipulable at will. Unfortunately, these inscriptions are often taken to *be* Nature and not products or interpretations of it. The sciences can, usually, be conceived in Latourian terms as carrying out sequential tests on actors, attempting to establish which "alliances" can or cannot be built between actors (Latour, 1993a, 1999). Note that these tests of "alliances" through trials of strength are something that Latour argues the researchers themselves are also put through – there is no ontological baseline where the testing stops for Latour. Thus when discussing a piece of scientific research, and the paper produced as a result of it, the authors of the paper are put through "terrible trials" by their peers (Latour, 1999, p. 124) in the review process in the same way that the researchers put their objects of study through similar trials in the laboratory.

Adding a further layer of complexity to this is the fact that researchers cross back and forth between said representations and the objects of study. This generates infinitely more interactions between humans and non-humans and is also the basis of Latour's view of 'interpretation'. Human meanings and organisations are applied to the non-human world, (and both 'sides' thereby *articulate* eachother through *Proposition*s) creating new alignments and entirely new worknets and "alliances" within the outer worknets. As is directly implied here – the act(s) of 'interpretation' itself are also objects of study, at least insofar as they leave clear 'traces' behind and have a definitive reality in their own right in that sense. This process – the use of Propositions, Articulation and instiutionalisation in order to capture, gather and redirect 'immutable mobiles' is the core of Latour's 'Circulating Reference'.

Criticism of Actor-Network Theory and Latour

There are some important criticisms of both ANT and Latour that warrant airing. However, a distinction should be made between the two targets. Whilst ANT is important for contextualising Latour's contributions, it is not essential for the purposes of this thesis. A rejection of ANT as a methodology (or, more correctly, a collection of methodologies) does not detract directly from the use of the Latourian concepts deployed here. However, as both are linked conceptually I shall address some of the criticisms of both consecutively.

ANT has a serious problem with amorality. The most telling critiques on this issue target the way ANT distributes responsibility via action. Whilst Latour's aim is to flatten ontological hierarchies, in normative terms some critics accuse him of excessively anthropomorphising objects by adding too many features of subjects to them (Collins and Yearly, 1992). Collins and Yearly point out that agency is effectively being distributed inappropriately to nature. Whilst humans are being objectified, nature is being anthropomorphised. It is partly in

56

response to this kind of concern that I argue in the next chamber that the useful parts of ANT and Latour should be considered in a more modular form. Furthermore, Latour admits to the amorality of ANT and explicitly rejects normative imperatives as an appropriate goal for it (Latour, 2005). There are some normative focused philosophies such as Floridi's (Floridi, 1999) that are compatible with Latour's focus on objects. As such I propose that such gaps are straightforwardly filled by any one of a possible selection of alternatives when required.

Latour does not regard this as a serious difficulty, especially as he would argue against a fundamental nature / human division in any case. However, should any questions of responsibility arise, ANT's cupboard is somewhat bare. (Smith, 2003) points out that associations and action is essential 'de-localised' by ANT. Attributing responsibility, or even to use Floridi's object focused and non-intentional alternative, 'accountability' (Floridi, 2006), thereby becomes difficult at best. If action is essentially 'alocal' as well as ANT being amoral, there's a potential description problem. I think this has force, and is a *serious* problem when talking about a highly dynamic and fast moving sociotechnical system. Latour's approach needs updating with a method that can cope with speed and also act as an effective 'gathering' mechanism for re-localising action in an intelligible way. For this I propose 'Computational Coherentism' (see section later in this chapter on Coherentism) as another compatible approach that slots into the gaps in ANT and Latour much in the same way as Floridi's contributions could fill the 'amorality gap'.

The expressly philosophical angles of Latour's work however, have a steeper hill to climb in terms of criticism. Harman, whilst often a sympathetic critic of Latour, also provides one of his most challenging philosophical obstacles. Consistent with his critique of the subject/object division, Latour's philosophy effectively has no 'philosophical atom'. Harman, in arguing for his own view of 'Object Oriented Philosophy' suggests that this is a problem.

For Harman and the other 'Speculative Realists' it is important to actually enumerate the 'big O' Ontology of what 'ultimate reality' really consists in (Bryant, Srnicek, Nick and Harman 2011; Harman 2011a; Harman 2011b; Meillassoux 2010). Even philosophies such as Floridi's that are highly compatible with Latour's have this 'atom' – in Floridi's case, 'information' is the universe's 'atom' from which all other entities derive (Floridi, 1999,2011).

The position of the Speculative Realists is that objects are effectively 'real but with hidden depths'. Harman claims that Latour's account does not allow for change as everything is exhausted in the relationships objects and humans have with one another (Harman 2011a; Kimbell 2013). Moreover, Latour's philosophy engages in a kind of 'occasionalism'. Harman argues that whilst it neatly avoids the subject/object gap, it does so at the expense of creating many much smaller gaps between actors and actants (Harman, 2009). Harman's critique has force, and makes sense if considering Latour's approach in more static terms. Reading any one complete Actor-Network account[18] does seem, after the fact, to have exhausted all of the possible actions that actors/actants could have taken.

This is, from my reading of Latour, one of the primary issues that beset anyone attempting to provide ontologically and epistemologically sound descriptions. Committing something to the written word is necessarily backward looking and post-hoc. I do not agree that with Harman that Latour's chosen ontology however also commits *anyone* using it to some kind of static metaphysics. Establishing relationships in situ is just a temporary stopping point down that road as Latour makes clear. Substance and endurance is provided for in Latour through both entailing continual work of Circulating Reference and the practice of Institutionalisation for maintaining worknets. Harman thinks he can dictate the terms of reality, in the process

---

[18] Such as (Latour, 1993a)

accusing Latour of causing an "absurd" inflation of entities that "balloon the scope of reality to an absurd degree" (Harman, 2011b, 2011a).

I do not think this is correct given Latour's careful observations regarding how certain scientific discoveries are operationalised. For example, once Pasteur's work had been widely accepted in France and beyond, the range and number of entities in the world was naturally multiplied by the assertion of the existence of microorganisms (Latour, 1993a). However, Latour's concepts can certainly be helped and enhanced to become more useful for fully capturing mobility and dynamism. They were not, after all, formulated by Latour originally with the digital in mind (just as 'network' was not formulated originally with a computer network in mind). It is one of the intended contributions of this thesis to expand Latour's work appropriately in just this way. With this expansion in mind I will now turn to coverage of the literature and debates in 'Coherentism' and 'Computational Coherentism'.

Coherentism

Coherentism is a relative newcomer in Western philosophy, only first appearing in clear accounts with F.H. Bradley (1914) and Bernard Bosanquet (1920). It is often represented by its critics as primarily a reaction to foundationalism. Both approaches, as classically understood, were largely responses to potentially devastating sceptical epistemological challenges.

The 'radical sceptic' provides arguments that seem intuitively plausible and yet provide a conclusion that we find completely implausible, if not repellant. The specific horn of the 'radical sceptic' position that both coherentism and foundationalism seeks to address concerns about is radical doubt regarding the foundation, the *justification* for any held beliefs. In reasoning about knowledge and beliefs, so the argument goes, we intuitively assume that

in having such beliefs we should also have good reasons to support them. Such reasons then beg further reasons for holding them.

This leads to what (Williams, 1999) refers to as "Agrippa's Trilemma" (Agrippa having been one of the earliest known philosophers to articulate this argument). If the sceptic is persistent in asking for reason followed by reason for one's epistemological justification, we are left with three alternatives:

   i)   An infinite regress of justifications or reasons to support each belief or reason expressed (the Regression Problem).

   ii)   A basic, foundational and yet dogmatic assumption or assertion beyond which we are unwilling to provide further reasons.

   iii)   Repetition of a previously stated belief or reason, leading to a circular argument.

To summarise the foundationalist and coherentist responses to the above in crude terms, foundationalism recognises the force of i) and iii) but does not recognise the force of ii). Coherentism recognises the force of i) and ii) but does not recognise the force of iii).

Foundationalism takes many forms – their differences to be found in varying accounts of what the 'basic beliefs' are and what it even means to be 'basic'. What they do have in common though is that idea of some kind of privileged class of 'basic beliefs' that provide rational justification for further beliefs. Mathematics and immediate experience figure largely as such examples. For foundationalists the regress of i) is stopped and the possibility of iii) is made irrelevant by this approach.

Coherentism on the other hand (also in numerous guises) makes some account or other based on arguing that some species of circular belief is not inherently bad epistemologically. Defenders of such theories point out that we do not just hold a random collection of unrelated

beliefs. Rather, beliefs tend to cluster systematically and here is the key premise – give one another *mutual* support. This has a significant intuitive appeal. However, that may be because it matches in both practical and ontological terms common shared experiences and perceptions of the activity of holding beliefs and obtaining or dismissing them. This appeal does not immediately lend epistemological weight (something that is done better in a related fashion by the 'contextualism' of Williams (1996) for example), so the other challenges to coherentism outlined below still need to be answered. The notion of mutual support however, does recall Latour's ontological position (Latour, 1993a) – that entities are made more real through having more links with other entities.

Laurence Bonjour is probably the most important critic of Coherentism. His account is of particular interest because he originally defended his own version of coherentism in a comprehensive work – *The Structure of Empirical Knowledge* (1985). He subsequently recanted and became one of coherentism's most distinguished critics, often referring back to his own original defences of the position[19].

Bonjour opens one of his critiques of coherentism with this assumption:

> I will assume that an *internalist* rather than an *externalist* conception of epistemic justifcation is correct, i.e., roughly that a belief's being epistemically justified requires that the believer in question have in his cognitive possession or be suitably aware of such a reason or adequate basis for thinking that the belief is true. (Bonjour, 1999, p. 118)

---

[19] I shall rehearse some of his arguments here as they effectively represent the 'best in class' representatives of such critiques and also highlight a number of the conceptual issues that are repeatedly encountered throughout this thesis in using more traditional philosophical approaches to current complex issues such as networked assemblages.

Bonjour regards the relative success of his choice of epistemological strategies (foundationalism) to be based upon a form of foundationalist *internalism*, against the perceived failure of a coherentism also based upon internalism. This is because, for Bonjour, it is only meaningful to discuss an epistemological foundation on the basis of an interlocutor being able to present their reasons for a belief whether to themselves or another. Being able to point to something *external* for him, effectively ends the debate and renders (internalist) coherentism irrelevant - i.e. for Bonjour coherentism only makes sense in this way with an internalist focus. As he later states, "this assumption…is needed in order for the foundationalism-coherentism issue to be even worth discussing." (Bonjour, 1999, p.118). This is the context of the 'philosopher in the armchair' being both the primary object of study as well as the instrument *for* said study. Note here also the redundancy of the internalism/externalism division for narratives acquired from a third party. The epistemological status of such an object as a third party narrative delivered online is not inherently a question of internal or external justification. Asking whether one already has justifications for believing in its veracity (internal), or can verify them elsewhere (external) does not advance our understanding of the situation. Hence the need to move onto ground that advances usefully in the direction of travel, whether that is through proposing computational, rather than human, substrates for internalist positions or coherence models based such a substrate in order to externalise the problem and make it amenable to outside information and changes[20].

Internalist foundationalism relies upon classical notions of *a priori* knowledge being possible,

---

[20] In this vein, Bonjour notes that one reason for the adoption of coherentism is the undesirability of the alternatives (in this case, varieties of foundationalism), echoing Harman's (2009) assertion that philosophers are rarely refuted, rather most are simply abandoned:
"as happens with rather alarming frequency in philosophy, the movement away from foundationalism in the last three decades or so often looks less like a reasoned dialectical progression than a fashionable stampede. And it is of course the rejection of internalist foundationalism that provides the main motivation for views like coherentism and externalism, neither of which could plausibly be claimed to be very attractive if it were not viewed by its proponents as the only remaining dialectical alternative" (Bonjour, 1999, p. 120)

or even a given, building epistemological justification quite literally from the inside out. I will not rehearse the standard objections to this here. It is of more relevance of this thesis to note that the 'stampede away from internalist foundationalism' as Bonjour characterises it (Bonjour, 1999, p. 120), is driven by a need closely related to the 'armchair' context noted above – the need to be self-contained. However, the problems we are addressing here are explicitly *not* self-contained. Everything bleeds across previously sharply defined categories in socio-technical interactions, making it unclear in many cases where any fundamental separations could be said to occur. Philosophers since the ancient Greeks have followed such a pattern of containment regardless – requiring what is in effect a hermetic seal where everything that is required is already present. Latour's philosophy directly challenges this on the grounds that every single thing present whether it is tangibly physical, like a molecule, or conceptual such as the idea of an armchair has a separate history and Circulating Reference chain, or chains or its own beyond the present place in space and time. Similarly, there are deep difficulties in accounting for how a human or bot agent in the digital wild can be said to 'know' anything in any kind of strong sense when particular narratives are always in play and often invisible. Expecting all of the necessary resources to be present is, from a Latourian metaphysical perspective, extremely naïve, though it is an unsurprising expectation if one is committed as Bonjour is to an internalist perspective, as are many philosophers who still to this day spend a significant amount of time attempting to appease the radical sceptic.

It is on such grounds, desiring a complete, self-contained account of the basis of epistemological justification, that Bonjour does not wholly reject the utility of coherence theories:

> While it is very plausible that coherence or something like it is *one* important ingredient in empirical justification, it is initially very implausible that it is the whole story (Bonjour, 1999, p. 122).

63

He thus appeals to the intuition that sets epistemologists down the coherentist path in the first place. Bonjour initially rejects it as the primary justification giving principle in spite of this for three reasons. Firstly, that he thinks a "pure coherence theory" entails no direct contact or input from the world beyond the belief system in question (including, possibly, other belief systems). Secondly, he argues that any system of beliefs could be made to cohere arbitrarily (e.g. by removing a key defeater for most of the beliefs in the system such as belief in the absence of evidence for a divine creator in a religious belief system). Finally, he also thinks on this ground that such systems could be created in an indefinite variety.

The version of coherentism I wish to defend is quite specific. I refer to it as *computational coherentism*, inspired by the work of Pollock and Thagard. Pollock does not necessarily consider himself a coherentist at least not on the common understanding of the term. And Thagard's particular account requires extensive exposition. Both taken together provide significant game changing contributions, respectively, for articulating epistemology in computational terms via the former and using the latter to resurrect coherentist models in epistemology. Furthermore, whilst Pollock does not take necessarily coherentism seriously in one of its more classical forms, I speculate that he may find a specific computational reimagining of coherentism based epistemology much more appealing.

The computational turn in Coherence theories is considered in the following chapter where I introduce the work of Pollock and Thagard as viable adjuncts to Latour's 'Circulating Reference'. Pollock provides an important, and fairly extensive, justification for re-adjusting the expectations and practices of philosophy through the use of computing. He champions the conjoining of epistemology and computer technology in order to create 'computational epistemologies' and 'remove the philosopher from the armchair' (Pollock, 1999). Against this background, Thagard provides a computational epistemology based upon a resurrected version of coherentism (Thagard, 2000). Thagard argues that many of the objections, of the

type that Bonjour raises, are answered when coherentism is reimagined in a computing substrate.

An immediate flaw may seem apparent here: Many models may include elements that would be expected to have multiple relationships with other elements in order to represent the state of affairs in question. This is especially the case if we are inputting instances of elastic narrative that are based on Circulating Reference and sundry actor-network concepts which by their very nature rely upon multiple relationships and have a core ontological dimension attached to the number of relationships involved[21]. In practice however this is a trivial concern as multiple relationships can still be modelled in a binary fashion[22] – the influence of multiple constraints is recognised by seeking overall maximal coherence.

A more serious problem, though it is one of practical difficulty rather than a fatal objection, is defining the elements and constraints. As Thagard puts it:

> To show that a given problem is a coherence problem…. it is necessary to specify the elements and constraints, provide an interpretation of acceptance and rejection, and show that solutions to the given problem do in fact involve satisfaction of the specified constraints (Thagard, 2000, p. 20).

Furthermore:

> Although much of human thinking can be described in terms of coherence, I do not mean to suggest that cognition is one big coherence problem. For example, the formation of elements such as propositions and concepts and the construction of constraint relations between elements depend on processes to which coherence is only

---

[21] i.e. the more connections, the more 'real' it is.
[22] As indeed, are all relationships expressed in a computing substrate as every machine's process could in theory be modelled on a single Universal Turing Machine given sufficient time and memory.

indirectly relevant (Thagard, 2000, p. 24).

He gives some clear guidance on how to characterise these elements and relationships appropriately to the kind of coherence sought[23]. Moreover, these specifications are used in his numerous algorithms to tackle each type.

Thagard contends that many of the failings identified in a coherentist approach are a result of human limitations and do not take account of the capabilities afforded humans now through cognitive outsourcing to machine agents. For the final part of this literature review, I now consider some of the simpler agents that are effectively used for such cognitive outsourcing.

---

[23] The practical specifications for different mode(l)s of coherence are very useful. However, it is not clear how they would practically interrelate in such theoretical approaches, even if such an interrelation does in fact resemble how humans might reason. Thagard responds by pointing to the interconnections that already exist between the various modes such as analogical coherence appearing in several even though analogical coherence also has its own specified mode worked out. This may be an effective rejoinder on a philosophical level yet on a practical computational basis it isn't sufficient, to my mind. Thagard simply asserts that new kinds of constraint could be included in the model of a coherence problem to represent constraints between elements that stand in for different types of coherence (in the philosophy of mind for example, with regard to the problem of other minds Thagard points out that we tend to reason about this in both explicitly analogical and explanatory modes). Whilst I agree with this in principle, in practice it is something that still requires further in depth research. Following his (2000) *Coherence in Thought and Action*, he has indeed explored the interrelation of different modes of coherence though the focus has been on emotional coherence, politics and ethics - e.g. Thagard's (2006) *Hot Thought*, which whilst of interest is not directly relevant to the epistemological concerns here.

A Taxonomy of Bots

To assist in understanding some of the references later in the thesis this last section of the chapter provides a generic 'taxonomy of bots' and a discussion of some of the relevant literature in such classification. It might be seen to imply that a general account of bots and networks, fused together into assemblages with humans, is possible *a priori*. This may have been a workable prospect if not for the fact that dynamism is an inherent characteristic of these worknets. The *work* in 'worknets' should always be borne in mind. This is why an empirically focused account is desirable where possible, focusing on specific instances. Such accounts of sociotechnical systems are replete in the broader software studies related literature, including Rogers' web interventions (Rogers, 2006), Montfort et al's analysis of a single line of code from multiple angles (Montfort et al., 2013) or the explanations of thinking behind programming in (Oram and Wilson, 2007). I also consider several case studies in chapters 4,5 and 6. These digital socio-technical systems have substantially greater materiality (ironic given their digital and virtual substrate) than many of the assemblages that have been considered in other ANT case studies such as Latour's account of researchers working with the flora and fauna of the rainforest (Latour, 1999) or the impact of Pasteur's discoveries on wider French society (Latour, 1993a). A formalised and *a priori* account would also lend itself to a simple technical account of the technologies in question whilst barely touching the more interesting philosophical, social scientific and policy ramifications of bot technologies and networked assemblages[24].

---

[24] The reader could be directed, for example, to the source code for a bot and then the formal specifications for every protocol and command used. Whilst informative and providing useful context, this would nevertheless only provide limited information on what this assemblage of code and protocols may do in isolation. Without the context of the larger assemblage, or worknet within which it is placed the exercise becomes somewhat navel gazing and of course, static. The 2010 Flash Crash amply demonstrated this principle with the objectives of numerous algorithms colliding to cause harm that would not necessarily have been apparent from a static analysis of their individual source code. A description of empirical examples where possible is therefore desirable.

Narratives in popular interdisciplinary works on ICTs tend to rest on a characterisation of the internet as the archetypal 'disruptive technology' reflected for example in (Shirky 2008; Zittrain 2009; Rheingold 2003; Tapscott and Williams 2008) and others. Even though there are some texts on how to construct bots (Hemenway and Calishain, 2004) or social media (Fitzgerald, 2008), or establishing a networked account of society (Castells, 2000) and a correspondingly enormous literature exists for computer and robot ethics, relatively few examples exist that exclusively speak to the activity of bots and networks. This is especially the case with regards to tracing the activity of Black Boxing and how to follow the deployment of such unprecendented speed and dynamism using contemporary means.

The first widely deployed 'bots' that had any kind of specific regular configuration identifiable in an online environment were used in IRC (Internet Relay Chat). They had names like 'NickServ' or 'MsgServ' – one to ensure nicknames in a chat room were not duplicated and the other to send asynchronous messages. They would often appear alongside human users in the list of names present in the chatroom and could be queried and commanded to carry out numerous operations. Malicious IRC bots began to appear around 1993, carrying out the very first intentional 'Denial of Service' attacks against hosts (Canavan, 2005).

The popular use of 'bot' can be traced to these IRC bots. Unfortunately, the recent use of the term 'Botnet' has been somewhat unhelpful and there are many completely contradictory characterisations of what 'bot' means, from classifying all malware as 'bots' (Holz, 2005) through to ignoring all uses of 'bot' that don't apply to website interactions (Schrenk, 2012). The literature in computer science surrounding the various varieties of malware is enormous and highly sophisticated, often pursuing mathematical heuristics to identify both source code typologies ("definitions") and specific patterns of network activity that may be indicative of

malevolent 'bots'. (Cohen 1990; Williamson and Léveillé 2003; Canavan 2005; Chen 2007) are examples, to name just a few.

The taxonomy below is by no means meant to be definitive or objective, rather it is an exercise in clearly distinguishing what the research here refers to and what it does not. It is not exhaustive and there is some bleed over between the classifications. The main distinctions are by the overall resulting behaviour rather than bearing strictly to whatever may have been intended by the original code writers. Following the cybernetics observation and notion of circularity, how exactly to classify a particular 'bot' or collection of 'bots' acting dynamically together is significantly dependant on the interactions with the environment. Numerous examples of this ambiguity are considered in subsequent chapters.

Firstly, there are three broad categories:

Malware

Synonyms / Subtypes: viruses, worms, spambots, spyware, rootkits

This category is a parking space for the kind of 'bots', some incredibly sophisticated, that are the primary concern for computer scientists and security specialists. The aims of their creators can vary significantly, however the intended end result is always malevolent, whether it is to wreak general havoc by slowing down or wiping hard drives or maintaining a passive 'back door' into a system that may never actually be used in practice. Furthermore, they concern primarily *technical* objects of study, and as noted above the literature and research into them is already very well advanced. That being said, there are some occasional *sociotechnical* aspects that are of note, for example, the vectors of infection are different when considering worms and viruses – the latter much more often involve a human component (forwarding emails, sharing memory sticks etc).

One aspect of the literature concerning these types of 'bots' that is of interest and inspiration to the current research though is the adoption of heuristic approaches. Rather than striving for exact description and detection methods, the application of the heuristic is a general means of capturing and extremely fluid and fast moving target. This thesis examines and proposes heuristic approaches for *sociotechnical* systems.

Daemons

Synonyms / Subtypes: scripts (stand alone), processes, libraries, modules, add-ins.

Whilst less intimidating than the broad category of malware, this distinction still covers too much ground for a single thesis. 'Daemons' refers to the various discrete automated processes that are generally stand alone, or where interactive with other processes, tend to operate within a single machine or highly discrete limited network, rather than the 'wild' of the internet.

Webbots

Inspired by the – somewhat casual – characterisations by Schrenk (2007, 2012) and Steiner (2013), this category covers the relatively simple (though with potential to be highly complex, especially in combination) automated scripts that are commonly deployed in order to interact explicitly with web sites or web resources. They still maintain the levels of speed and dynamism that place them within the range of networked assemblages that pose a problem for contemporary methodologies to describe and characterise that this thesis addresses. It is possible that limited conclusions can be drawn from their study outwards to the much more problematic flora and fauna of the previous two categories.

Instead of listing some subtypes or synonyms on a separate line as above it has been useful to delineate several varieties of these 'Webbots' that are deployed in the research expanded upon later. Their subtle differences in operation enable the teasing out of numerous nuances

of description that illuminate the overall project of general description of networked assemblages.

For clarity, following this chapter, 'bots' will be used to indicate the varieties of 'webbots' below, excluding the two categories already covered above unless explicitly stated otherwise.

Spiders

Synonyms / Subtypes: Web Spiders, Crawlers, Web walkers, Aggregators

These 'webbots' are designed explicitly to index and follow links, webpages and other resources that they encounter online. They can be designed to follow specific ranges or left running to cover substantial tranches of the internet. All of the major search engines deploy their own 'spiders' in order to continually refresh their databases of internet content.

Scrapers

Synonyms / Subtypes: Screen Scrapers

Lots of content available online is not necessarily accessible in desired formats. Sometimes machine readable data needs to be human readable and vice versa. Scrapers enable bespoke solutions to data discovery that may not be possible through already existing means – sites may offer various outputs, including human centred presentation layers, RSS feeds and even APIs (Application Programming Interfaces) and yet the content may not meet requirements. They may also often be used to copy data from a website en masse for subsequent processing. Sometimes their work is enhanced with *Mechanizers* (see next).

Mechanizers

Synonyms / subtypes: Snipers, Pokerbots

The name chosen for these 'webbots' is inspired by a Python module that can be used to

construct them, 'Mechanize'[25]. Out of the three sub-types of 'webbot' listed here these are the type that will most often attempt to pose as a human agent. They are used wherever a site has interactive sections that require clicking, selecting and so on. Many websites do not yield raw data sets, only presenting discrete selections chosen by web based interfaces (i.e. by date, type, keywords etc). Mechanizers are often used to collect the information contained in a back end database that would otherwise remain inaccessible. Subtypes include more explicit human-replacement functions such as so-called 'Snipers' that may make ebay bids on behalf of a human user as soon as a certain bid price is reached (these are much simpler versions).

Following this review of literature and a brief forary into the operational concepts and terms that will be used throughout the remainder of the theis, I now turn to a much more in depth consideration of the theoretical computational and digital concepts that figure most prominently – Latour's 'Circulaing Reference' and 'Computational Coherence' inspired by Paul Thagard.

---

[25] See the Sourceforge page for technical details http://wwwsearch.sourceforge.net/mechanize/

Chapter 3: Theorising the Computational and Digital

If how we apprehend the complexity and speed of networked assemblages is problematic, then one of the key aspects of this is how we explicitly *refer* to them. That is to say, the relationship of *referent* and *reference* needs to be established. Fixing, or even making sense of, references can be problematic with static objects. Dynamic and fast moving referents provide an additional challenge. I propose that Latour's concept of 'Circulating Reference' is sufficiently adaptable when applied alongside 'Computational Coherence' to this end, providing a viable means of addressing how to *reference* such fluid objects, entities and events and choose between competing accounts of digital events respectively.

The following account, of necessity, moves from the abstract to more concrete, starting with an outline of Latour's metaphysics and ontology, then moving onto Circulating Reference proper and secondary (but important) related concepts including Whitehead's 'Propositions'. Finally, the contribution of these notions to the general problem of apprehending networked assemblages is considered before progressing to a consideration of Coherentism.

There are no simple *static* solutions to dynamic problems situated within (or based upon) fast moving sociotechnical networks. What is clear is that, both in epistemological and practical terms, the patterns described form delicate relationships that are potentially easily interrupted or disrupted. In many cases the technologies themselves, and the practices built around them, quickly adapt to such changes (such as the oft used cliché regarding network traffic 're-routing around damaged nodes'). Describing such states of affairs using typically static notions of sense and reference such as those proposed by Ayer (1990), Russell (2001) and so on is inhibited from the start if referent and reference are in regular (if not always continual) flux. What is needed instead is an approach that contains within it a distinct elasticity – it can move, change, be bent out of shape and still usefully circumscribe the dynamic object(s) of

study. Latour's Circulating Reference is able to meet this need. The idea is largely underutilised even by Latourian and ANT inspired scholars despite having great utility as a method for providing *elastic narratives*. It also requires some adaptation to digital and networked objects. When Latour first explicitly introduced the idea in *Pandora's Hope* (Latour, 1999), he was applying it to worknets that moved at a relatively modest speed. His primary case study was the research process of scientists examining the flora and fauna of a small part of the Amazon basin. He used the concept to explain and trace how the soil samples and other findings 'on the ground' related to one of the final end results – academic papers. The worknets of networked assemblages are significantly more frenetic than the potential months or years long process of producing academic papers from soil samples.

By their very nature, references in the kind of sociotechnical worknets of concern here are highly mobile. Words can refer to objects in continuous flux, such as a database that is in accessed and updated several times per second. It is manifestly not the same numerically identical object from one moment to the next. It can change rapidly in size, content, accessibility and so on. In the case of newer database technologies such as 'Cassandra'[26], designed to tackle Big Data and cloud computing issues, it can even dynamically 'shard', i.e. split itself into smaller components. On the other hand, multiple objects or even entire assemblages can slip rapidly in and out, with a single word or reference referring to whatever fills that space. A typical connection to a website across the internet is like this. When one refers to the ongoing 'connection', one is in fact referring to packets of information being exchanged back and forth across a population of entities that may change from one second to the next. Until it breaks (and the Black Box of 'connection' is essentially forced open), this occurs invisibly to the user, whether human or bot.

---

[26] The specification of this database technology can be found here: https://cassandra.apache.org/

Whilst many of the singular entities involved (bot, web user, webmaster etc.) can usually be classified with relative ease, the particular happenstance of their interactions, or rather the particular *pattern* of their interactions requires a more dynamic approach to capture. And, more than just dynamic, also with a high toleration for error. This means using a method that refers to a *range* rather than just a static object or value. Latour's underlying metaphysics has the advantage of an epistemological inheritance that is natively object friendly in the sense that *objects are taken seriously*.

Latour's Metaphysics and Ontology

Latour's unusual metaphysical views, at least as they were expressed in 1986, are primarily matters of ontology. Harman (2007) refers to it as "relational ontology" – conferring reality and distinctiveness through relations. Latour's philosophy defies straightforward categorisation as Realist or Constructivist. Latour's approach embraces both. And to understand why, one must first understand his metaphysical position. Harman places the blame for this confusion on the continuing stale division of Western philosophy into continental and analytic schools. Efforts continue to build bridges between the two, yet Harman argues, there is no need for such rapprochement. Both have an "unspoken allegiance" to what Latour called Kant's "Copernican Revolution". Kant made solid the rift already in various forms inherited from previous philosophers between subject and object – between human mind and the world. Both Latour and Harman consider this unacceptable, for a "situation has resulted in which language is regarded as a central philosophical problem, but not the relation between tires and concrete" (Harman, 2009, p. 32).

Latour effectively steps over this chasm between subject/object in an ingenious way. His metaphysical view is primarily one of ontological flatness and – importantly - *bi-directionality*. In *Irreductions* Latour states that he opposes "the assumption that force is

different in kind to reason; [that] right can never be reduced to might" (Latour, 1993a, p. 153). Harman is at pains to indicate that initial readings of such statements may mistakenly cast Latour into the strong social constructivist camp where power is the dominant phenomenon. Conscious of being misunderstood in this way, Latour cycles through many examples and refinements of his position in several of his works - most notably in (Latour, 1993b, 2005). As Harman points out, in Latour's philosophy, might is just as reducible to right as vice versa. They both, potentially, exert equivalent levels of force in the world.

One must tread with care when attempting to provide an exposition of Latour's underlying philosophy. There is much discussion of 'objects' and 'networks'. However, it is not necessarily clear from isolated quotes that the two are completely interdependent and indeed define one another in his account. Equally, whenever Latour is referring to 'actors' or 'actants', he is usually referring to objects in one sense or another and the networks and power plays in which they enmeshed is never far from view. Objects, in his view, are interchangeable concepts with Black Box and Actor-Network. The different words are primarily matters of emphasis. A completed exchange between a bot and a website, for example, can seem like a Black Box from the outside. Even if I have programmed a bot, when launching it to carry out its task I'm not likely to be sat in front of the monitor, watching it carry out its calculations (unless something has gone wrong). I simply wait for the output to arrive – which could be a line of text, a file, or many other things. This is despite the fact that an immense amount of complex action is taking place, hidden from view, yet easily accessible if desired. A temporary assemblage[27] has been formed beneath the surface then dissipated beneath my awareness.

---

[27] Indeed, a temporary *Black Box*. Just because it is not witnessed by a human does not stop it functioning as a Black Box for non-human actors.

Every object – and to properly conceive of what Latour intends here, I will use *actor* or *actant* to describe this further because of the baggage that comes with 'object' – exists on a level ontological playing field. For Latour it has been a colossal mistake for philosophers to attempt to line up fundamental ontological hierarchies and categories that were said to fully capture aspects of the universe or the division of entities within them. If there is anything fundamental at all about the universe for Latour, on a capital 'O' Ontological understanding, it is relationships and associations. One of the most difficult aspects to grasp here is Latour includes many actors or actants that would usually be written off as completely unreal and undeserving of formal philosophical or existential status. Concepts and fictions such as 'force' or 'Mickey Mouse' respectively are on the same ontological level as a car or a hammer. This is the case whether one endures much longer than another. Why? Because they exert influence – they *act* (actors) and are *acted upon* (actants). This is why it is more intelligible to, at least temporarily, drop references to 'objects', even though in Latour's lexicon they are essentially synonymous. The fact that this approach provides the ability to account for force or influence applied by *fictional* entities lends Latour's methods to similarly accounting for *virtual* entities too.

This move is incredibly profound, in philosophical terms because, as Harman puts it:

> It…ends the tear-jerking modern rift between the thinking human subject and the unknowable outside world, since for Latour the isolated Kantian human is no more and no less an actor than are windmills, sunflowers, propane tanks, and Thailand (Harman, 2009, p. 14).

Moreover, for Latour, *contra* Aristotle and his inheritors, an actor does not have a hidden essence, covered or hidden with a surface of accidental features. It is fully deployed in the world. It is completely concrete. That we only ever grasp aspects of it isn't anything to do

with some kind of radical relativism about ultimate reality, it is because we only perceive it and articulate it through the medium of other actors and all of the relationships between us, including via the process of translation. This is what he means by irreducible. We *do not reduce* one actor into another when we explain or describe it. We always explain it instead in terms of other things – other actors, actants, objects and relationships. *We articulate actors in terms of other actors*. This makes Latour's approach highly congruent with Coherentist perspectives. I can 'explain' a bot in terms of its relations with other entities (human agents, servers etc). I can also 'explain' it in terms of its *internal* relationships – the source code and the resources each line of code or invocation of other code libraries draws on. Both are valid and equally 'real' ways to grasp the nettle of description. Both are *articulations*, and in combination that add layers to the 'reality' of the actor. The actor is *more articulated*, in Latourian terminology.

Such explanations are convincing in so far as one is willing to carry out the proper labour, as Latour sees it, to situate and transform the target actor(s) in terms of others. It always takes *work*. I argue that this notion of both initial and continual effort is probably the single most important aspect of Latour's approach. There is a *price* for every description or explanation. The chain of relationships, equivalences and translations always has a cost and it always *risks failure*. This is the importance, in (Latour, 2005) of reconsidering the use of 'Actor-Network', or 'Network' and replacing it with 'Worknet' instead as discussed in Chapter 1. My own attempt to articulate relationships between press releases and online media outlets initially failed in just this way as mentioned in the introduction (see also Chapters 4 and 5) – as did similar informational relationships that permeated the Flash Crash (see Appendix 1), and the power outages in Ohio in 2003. My work was passing through the *obligatory passage point* of the Churnalism facility, which suddenly frustrated my efforts when its own internal composition was altered by its developer. Similarly, when the circuit breakers were

switched on (or came on automatically), during the 2010 Flash Crash, they prevented traders, both human and machine, from accessing the obligatory passage points of the markets and market links they were expecting to be able to access (see Appendix 1). Moreover, one of the most substantial aggravating factors in play was that the circuit breakers were not applied across markets in unison, causing chaotic re-routing of orders and equally chaotic informational assessments as to the current state of play.

At this point the notion of *mediator* becomes useful in understanding what Latour means. If we can and do only understand actors and networks in terms of other actors and networks, it is correct to point out that none of these points provides a completely transparent intermediary. There is always a cost for passing through, however small (or large). Some kind of charge (or service in kind) is levied for use of electronic trading venues for example. However even at the raw, physical level digital signals may seem exact, transparent and complete (discrete) yet the cable or radio transmission that carries them in a physical medium requires energy and the exact waveform that carries them is never exactly the same twice. There is a cost to the transmission and the *physical* signal is always ever so slightly altered or changed in the process. The router, the cable, the radio wave, all act as mediators and all exact a cost and a tiny transformation of whatever passes through them. Every actor exerts some kind of force or influence, and it is resisted by a great many other actors.

The primary distinguishing feature between actors in terms of what they are able to achieve is not necessarily something inherent in them. Rather, it is to be understood as relative differences resulting from having more or less *alliances* with other actors, thereby providing a relative measure of strength or weakness. Alliances here referring to the way actors enhance or resist one another is in highly specific ways – and it is the analysis of this that provides the core of Latour's empirical focus. Each actor can be found at a specific time and place in the world and with specific alliances. In some accounts of empirical phenomenon provided by

Latour and others inspired by his thinking such as (Berry 2011; Rogers 2006; Venturini 2010), one of the simplest ways to express these phenomena and the practice of their study is in attempting to describe *trials of strength* between them. More associations and allies means greater strength, being cut off from allies or associations means weakness.

Such is the account Latour gives in *The Pasteurization of France* in detailing how Pasteur's work and influence came to dominate his contemporaries in the nineteenth century. One can see a strikingly useful analogy in the operation of cloud computing – incoming demand is satisfied *elastically* by recruiting more servers (allies) whenever demand goes up. If one wanted to attack and incapacitate such a server farm for cyberwarfare purposes, one would have to bring an equally dynamic and powerful set of allies to the interaction. And of course this is now typically the case when such events occur. As described in broad surveys of the state of the art such as (Carr, 2011), botnets are now comparably vast, using tens of thousands of separate computers as 'zombies' in order to attack resilient networks such as cloud based computing networks. This is in comparison to previous decades where a single computer could be powerful enough to attack a third party server, which was unable to call upon additional resources (allies) to fend off the attack.

Latour simply steps across the abyss of subject/object and mind/world. He gives up on what Harman refers to as the "stale debates" both between analytic and continental schools of thought and deep concerns about whether the external world is constructed in the mind or has some independent existence of us. For him, there is a continual constructivism at work but it is not simply 'social' as it might be commonly understood. No, in the ongoing *trials of strength* between all actors, the constructivism is something that every actor participates in, not just those of flesh and blood with a conscious self-awareness.

There is also an important and measurable sense in which actors become more or less 'real' on this basis. The more relations an actor has, the more connected they are, in the world and the more it can leverage these relations to impose its own forces, the more 'real' it is, quite literally because it imposes on the wider world of actors (even if just very locally) more than its otherwise ontologically equal peers. Each relation is an articulation, and often an articulation that occurs in both directions. Botnets, search engines and algorithmic trading rely heavily on this kind of ratchet effect. A bot, or a human deploying bots, can exert influence at multiple locations simultaneously – speed, positioning and ubiquity of agents thus becomes crucial in exerting virtual power. Being able to initiate and maintain anything from dozens through to thousands of relationships with other potential counterparties to an interaction really does make the power of that bot or human more *real*. Thousands of such digital traces can lead to extreme drama, ignominy and losses in the "meatworld" of humans. Ranging from the online presence of a business being taken down to a swathe of nodes in the power grid ceasing to function.

Truth in an important sense here actually becomes synonymous with what is describable in these terms. It is a *result* rather than a starting point. As Latour puts it:

> A sentence does not hold together because it is true, but *because it holds together* we say that it is 'true.' What does it hold onto? Many things. Why? Because it has tied its fate to everything at hand that is more solid than itself. As a result, no one can shake it loose without shaking everything else (Latour, 1993a, p.185-6, original emphasis).

For me this clearly allies Latour with a Coherentist approach, prefiguring much of the discussion to be covered in the second part of this chapter and also places him in a similarly object-oriented context allowing for such relative definition between objects being one of the

unifying themes of broadly conceived 'object oriented philosophy' tailored to dealing with the highly dynamic networked assemblages as discussed in Chapter 6.

There is thus a measurable, graspable, observable cost or work associated with establishing the properties of an actor, as they occur and act within the context of (and often at the expense of) other actors. Latour explains this kind of approach in terms of "relative relativism" as opposed to "absolute relativism" as he puts it in *We Have Never Been Modern* (Latour, 1993b, p. 111). This is an important caveat to grasp – *all* actors resist one another, it isn't as the Kantians suppose, a matter of purely human significance and cognition. It is abundantly clear from Latour's empirical examinations into specific science and research practices that he thinks that the world very much *resists* human fabrication. Our understanding of many phenomena are highly socially constructed but that does not escape from the underlying realism of Latour's account that *something pushes back* and the relativism comes not from the "nature" of ultimate reality but from our own limited (and varied) means for observation (via other objects/actors) and the very networks that we find ourselves in. For all of their power to affect human affairs, networked computers and bots are soon stopped in their tracks if the power is cut. Anything *does not* go! As Latour argues, in response to 'absolute relativists'

> Discourses and associations are not equivalent, because allies and arguments are enlisted precisely so that one association will be stronger than another. If all discourse appears to be equivalent, if there seem to be 'language games' and nothing more, then someone has been unconvincing. This is the weak point of the [absolute] relativists…By repeating 'anything goes' they miss the work that generates inequivalence and asymmetry. (Latour, 1993a, p. 168-9).

There is that all important notion again – *work*. It requires tangible effort to initiate and maintain any state of affairs. One of the biggest problems with static conceptions of reference is the absence of such an activity based notion. A reference between word and world either holds or it doesn't. There's no room for error, however there's also no room for usefully grasping the messy, contentious entities with which we customarily engage in the world (not least of which are eachother, as human beings interacting). If there are 'brute facts', they exist within a context of effort – the world 'out there' isn't just there – *it pushes back*. Effort is also required to study and describe them, not to mention to maintain such understandings once they have been grasped. Even apparently stable entities or phenomena have a history of work (something Latour variously refers to as a 'trajectory' or the process of Institutionalisation). Mind bogglingly complex, long and hard fought for in many cases. Take the form of any breed of dog, for example. A combination of both millions of years of evolution and artificial selection through breeding has produced any one dog one might encounter. To the extent that any particular breed may appear to stay in a relative state of similitude it is primarily because humans keep them that way in both individual condition and in evolutionary prospects for their descendants. One can see this work at light speed in the interactions that take place in genetic and evolutionary algorithms, where thousands of iterations and interactions are used to optimise code for a specific purpose. It may be difficult to grasp, especially in so short a space of time, however every program ever written has a history of work behind it and does not, like any object or assemblage, simply spring forth *sui generis*. Black Boxes are treated as solid entities only in so far as they continue to *appear* to be solid and uncontroversial. We're able to suspend consideration of the colossal chain of networks, circumstances and actors that brought any one Black Box into being.

It is through the use of the Black Box that Latour is able to reintroduce some kind of practical

notion of (relative) hierarchy in perceived reality and empirical observation – a hierarchy lost by placing everything human or non-human, fictional or non-fictional on a level ontological plain. Black Boxes also act as relative levels of containment. The computer that printed or displayed this document functions much as singular Black Box. Yet it consists of a great many Black Boxes – both real and virtual, any of which can come to the fore, cause the machine to malfunction (or underperform) and remind us that the computer is actually a controversial linking together of countless Black Boxes. And moreover, it is a Black Box of Black Boxes that can change its relative virtual constituent parts on a regular basis. Any Black Box can be rearranged and challenged.

Black Boxes can be seen at work in purely conceptual terms as well as physical (or hybrid) constructions. As Latour goes into in some detail in *Science in Action*, scientific theories go through many blistering trials of strength. They only become Black Boxes when there is no longer any reference to origins, controversies, struggles or even the primary theorists responsible. As he puts it, "who refers to Lavoisier's paper when writing the formula H2O for water?" (Latour, 1987, p. 43). The Black Box is a gathering of actors and relationships that went through a torrid phase of intense struggle and subsequently becomes something relatively settled, with minimal maintenance requirements. At least until something goes wrong and one of its constituent objects/Black Boxes reasserts itself.

Another hierarchy building aspect to this concept of Black Boxes is how they become obligatory passage points. We are surrounded by countless actors that we customarily ignore, do not interact with or remain unaware of interaction. Many Black Boxes exist in typical human life however that force other actors to pass through them or be subject to their influence

Black Boxes in this context remain neither socially generated phantasms nor solid concrete substances. Rather, they occupy the space of *performance*. An ongoing one that requires

effort to maintain, just like their constituent objects and their observation. To the extent that they maintain solidity uncontroversially it is because we, passively or actively, maintain a list of *trials of strength* to which they must be constantly subject; thereby providing the empirical grounds by which we account for them, use them, deploy them and describe them. This list is "the beginning of its definition" for Latour (Latour, 1987, p. 89).

Circulating Reference

When discussing the scientific expedition he accompanied, Latour says of the point where the team were comparing two maps:

> Remove both maps, confuse cartographic conventions, erase the tens of thousands of hours invested in Radambrasil's atlas, interfere with the radar of planes, and our four scientists would be lost in the landscape and obliged once more to begin all the work of exploration, reference marking, triangulation, and squaring performed by their hundreds of predecessors. Yes, scientists master the world, but only if the world comes to them in the form of two-dimensional, superposable, combinable inscriptions. (Latour, 1999, p. 29).

The scientists here are essentially using cognitive and epistemological shortcuts, developed through thousands of hours of labour that in some sense do actually relate consistently with the phenomenon they purport to describe. And they will continue to do so for as long as the 'chain of translations' remains largely intact[28]. The scientists are completely distanced from the phenomena they are pointing to and discussing. Yet they refer to features, and even a photograph of one area that "bears a mere resemblance, in certain traits, to figures printed on

---

[28] A devastating meteor impact on the Amazon basin for example would break the chain of translations dramatically – however it is reasonable for the scientists to assume some kind of continuity because, as a general rule, large geological features change only very slowly.

a map," with assurance (Latour, 1999, p. 30). The artefacts of maps (amongst many, many others) are at once completely unlike the phenomena they somehow represent or capture and yet have the capacity to render and communicate those phenomena with relative accuracy.

How does this process begin? Latour describes one of his first experiences in the field – noting that one of the scientists charts and numbers several hectares on a grid using Cartesian coordinates and

> These numbers will allow her to register the variations of growth and the emergence of species in her notebook. Each plant possesses what is called a reference, both in geometry (through the attribution of coordinates) and in the management of stock (through the affixing of specific numbers). (Latour, 1999, p. 32).

Software Studies accounts use a similar approach to outline how discrete digitisation is applied to break down and categorise external phenomenon (Fuller 2006; Berry 2011). Latour observes that this is *not* the creation of science *ex nihilo*. The scientists are already bringing with them enormous chains of translation in order to make sense of, perceive and categorise their object(s) of study. When a number is attached to part of the grid, it signifies that even in the midst of the forest, Latour and his companions are in a *laboratory* of sorts, even if as he notes – a minimalist one. One can also regard the computer as a similar kind of laboratory. One carries out this dividing, categorisation and labelling for example in creating a database for a music collection (or family and friends' details, pets, wildlife in the garden and so on). Moreover, it is usually created in a spreadsheet or 'dataframe' format – a grid. This particular grid-like form is thereby already transforming the observed matter into a form more similar to its end referent – marks and diagrams on paper. In terms of tracing the activity of scientific inquiry from the root, it appears almost contradictory: "I rediscover the

tautology that I believed I was escaping by coming into the field. One science always hides another" (Latour, 1999, p. 32).

A perfect (i.e. objective, value-free, *sui generis* or *a priori*) starting point is therefore impossible, and the desire for it quite likely an artefact of the dominating narratives in epistemology requiring both absolute certainty and a bridge over the (supposedly impassable) gap between words/self and world. One can only start in the midst of messy connections and translations already in progress, charting an epistemological structure that is less foundational and more *rhizome* like. Indeed, Latour notes that, "We always forget that the word 'reference' comes from the Latin referre, 'to bring back'" (Latour, 1999, p. 32). The gap between reference and referent does not seem quite so intimidating or insurmountable in this context; where it is possible to 'bring something back' because we are already, collectively and individually, enmeshed within it and not as so many philosophers following Kant have supposed fundamentally separated or distanced from it.

Similarly for the botanist accompanying Latour who is collecting samples of plants from the region, she cannot collect the entire population, so samples are preserved as references that can be consulted in the future. Each specimen is a "silent witness" and "guarantor" for future claims to be made about the investigations carried out during the period of their collection. The subsequent process through which the collected samples go is, Latour notes, much more heavily investigated and understood by science studies scholars. A shelving unit at a faraway institution is used to place samples in x, y coordinates. The piece(s) of furniture in question, themselves present a theory of classification, organisation and representation. How are we, our references and referents then located in such cases? A "tiny part" is used to grasp the "immense whole". It is a form of both displacement and delegation. The "tiny parts" (such as soil samples, leaves of a plant) are delegates for the larger whole and guarantors of a physical link.

'A tiny part grasping the immense whole' is an interesting phrase to describe the process at work here and hinting at the fact that there is a solid relation. Or rather, the attempt is made to forge one between referent and reference even though there are successive *transformations* with each intermediary. "Space becomes a table chart, the table chart becomes a cabinet, the cabinet becomes a concept, and the concept becomes an institution" (Latour, 1999, p. 36). The issue of what kind of distance is involved here seems less relevant than the idea that "we have transported a small number of pertinent features." This invariant (or semi-invariant) reference is the core of Latour's concept of Circulating Reference. And it is because of this invariant, and this movement from forest to laboratory (or rather, a more structured, formalised laboratory) means gaining more than is lost in the process.

Why? Latour lists numerous reasons: First, a kind of convenience – everything of use is gathered in one place. "The supposedly vast distance between writing and things is now only a few centimeters." (Latour, 1999, p. 38). Secondly, this gathering enables a conjunction and comparison that would not be possible in the wild/field. With regard to the plants gathered in the Amazon for example, one can immediately compare them, on the table, with samples obtained on the other side of the world, "…all visible under the same unifying gaze." Thirdly, the objects of study can be shifted, shuffled and recombined in novel ways, the "Plants are not exactly signs, yet they have become mobile and recombinable" (Latour, 1999, p. 38). If such mobility can be used as part of the description and reference to mobile objects of study in the case of plants then it makes *even more* sense to apply such a method to the fundamentally peripatetic socio-technical systems of concern to this thesis.

Such advantages enable the researcher to find patterns that were previously opaque as "Scattered through time and space, these leaves would never have met without her [the researcher] redistributing their traits into new combinations." And "In losing the forest, we win knowledge of it" (Latour, 1999, p. 38). Unfortunately, it is also possible for confusion to

return at any point – the chain of displacement and transformation can easily be broken as the data and samples rapidly pile up:

> We have barely arrived when we must leave; the first instrument is hardly operational when we must think of a second device to absorb what its predecessor has already inscribed. The pace must be accelerated if we are to avoid being overwhelmed by worlds of trees, plants, leaves, paper, texts. Knowledge derives from such *movements*, not from simple contemplation of the forest. (Latour, 1999, p. 39).

There is always a "risky intermediary pathway" between objects to words. Referring to the culled plant specimens, Latour notes, "The samples will remain attached to their original context solely by the fragile link of the numbers inscribed in black felt-tip pen on the little transparent bags" (Latour, 1999, p. 46). The activity of taking such pathways for Latour is fundamentally about transformation right from the very start: "For the world to become knowable, it must become a laboratory" (Latour, 1999, p. 43). Researchers are,

> Sent by institutions that are thousands of kilometers away, obliged at all costs to maintain the traceability of the data we produce with minimal deformation (while transforming them totally by ridding them of their local context). (Latour, 1999, p. 47).

A kind of *property swap* then occurs. The properties of the local context are swapped for mobility, traceability. And with it comes risk. With risk comes subjectivity, substantive relativism, relationships and *reality* ("reality flows back in"). The discipline required is enforced by the use of ingenious devices that are, in effect, very real hybrids between the world of 'things' and the world of 'signs'. One such device Latour highlights, used by the Amazon reseachers, is called a 'pedocomparator'. Samples already gathered from the makeshift laboratory of gridded regions can be placed in a similarly gridded rectangular

pattern in a portable device, enabling not only convenient transportation but also instant comparison of diverse soil samples. Such devices – maps, gridded hectares, pedocomparators, because of and/or *in spite* of their artificiality, enable us as humans to create backgrounds against which "phenomena will be able at last to appear, that is, to stand out against the new backgrounds we have astutely placed behind them" (Latour, 1999, p. 49). As soon as the material soil is either marked out by the artificial grid, or poured into the pedocomparator, it becomes a sign.

Latour argues that the philosophy of science and epistemology miss such crucial transformations and thus set up the 'abyss' to be crossed, unwittingly and artificially. Especially as it is usually only the resulting abstractions that are studied by these disciplines. This "moment of substitution" is *crucial*. It is the "very instant when the future sign is abstracted from the soil" (Latour, 1999, p. 49). We are not leaping inexplicably from the material thing to the idea of it, but through a *continuous*, and importantly, *traceable* series of steps. We are, to use Latour's sense of this word, hereby *articulating* the meaning of this piece of soil. Again, this characterisation is even more applicable to the hybrid devices of computing technology, these ever ubiquitous mobile laboratories (many of which we now carry on our persons) that cross the world of 'things' with the world of 'signs'. Within them are the inscriptions, millions of them, from programmers and hardware manufacturers that encode these multiple crossings between 'thing' and 'sign', providing representations of timezones, geography, energy levels, motion, geometry and on and on. The inhuman objects, these opaque Black Boxes, suddenly become visible components in a vast chain of motion linking referent to reference in an ongoing dance. A signal transmitted through multiple physical intermediaries.

Appreciating the motion and the ever so delicate transfer of meaning from one object to the next (say from the atomic clock, to the computer server, through the internet and then to your

timekeeping application/device) provides a whitening of the Black Box, a change in perspective that guards against complacent ignorance. To put it more clearly, to bring the Black Box(es) forward into conscious apprehension in a way that is useful to us, and to do so on a regular basis means to regard them as mobile, active and in some sense *living* (socialised) entities. It means lifting the statement 'the cat is on the mat' off of the page and adding a chain of observers, or delegated (non-human) observers, between you and the room where the cat may or may not still sit on the mat. If the cat has moved from the mat and there is a break in the chain of observations and communications, the Circulating Reference has failed. There is no doubt that it is a delicate thing indeed, albeit something strengthened if there are *multiple* chains of reference. Articulation and Circulating Reference are fundamentally similar and crucially related in this respect and cognate concepts for any kind of Coherentism, especially when considering multiple chains of Circulating Reference. The jump between 'thing' and 'sign' mirrors that between object and subject.

This supposedly vast conceptual gap is actually everywhere around us and so small in most cases that it is crossed with barely a thought. Applied mathematics for example, is one of the most common domains for this profound yet ubiquitous leap. Latour notes that surprise is often expressed at the efficacy of mathematics in that it can be consistently applied to the physical environment. This is because the 'leaps' it makes as a discipline and practice are not generally enormous jumps across chasms such as subject/object or nature/science[29]. Rather it makes tiny, sometimes infinitesimal, leaps between the measuring instrument and the object being measured. A ruler placed against a soil sample for example, already has a key likeness with the sample in that both are geometrical entities and the "tiny" gap crossed is simply that between the size of the object and the approximate measurement grasped via means of the

---

[29] This is not necessarily the case in the mathematics of high energy or subatomic mathematics where such tiny differences matter a great deal and likely contribute directly to many of the present controversies in these subjects.

ruler. In terms of the transformations described in the work of the Amazon scientists, it follows this kind of pattern:

> The prose of the final report speaks of a diagram, which summarizes the form displayed by the layout of the pedocomparator, which extracts, classifies, and codes the soil, which, in the end, is marked, ruled, and designated through the crisscrossing of coordinates. (Latour, 1999, p. 56).

Computers and computing technologies, then, cross this tiny abyss millions of times at lightning fast speeds and, moreover, do so generally without our awareness. This, in spite of the fact that there are an infinite number of gradations. True OR false is a synonym for the choice between subject OR object. Our measuring devices, our computers of all types, take the risk and continually leap across, delegated on our behalf to do so. Instead of 'truth' what we are more accurately doing when we use them as tools with which to observe and understand the world is *assessing risk*. Primarily this is assessment of the risk that the chain of reference is broken at some point. If this chain of transformations is sufficiently rigid[30] and unbroken then "We never detect the rupture between things and signs, and we never face the imposition of arbitrary and discrete signs on shapeless and continuous matter" (Latour, 1999, p. 56). The reliability of the reference is maintained as it circulates through each stage of successive abstraction, each directly traceable to the previous one. The actual act of reference in practice is not about seeking some external guarantor of its authenticity; rather it is about maintaining something *constant* through each transformation:

> Knowledge does not reflect a real external world that it resembles via mimesis, but rather a real interior world, the coherence and continuity of which it helps to ensure. (Latour, 1999, p. 58).

---

[30] This rigidity in communication, and the lengths (of the chains) possible, is one of the primary strengths of the current sciences, for Latour.

Note Latour's explicit reference to internal coherence. It is also important to note that the 'interior world' to which Latour refers here is not that of the epistemologist's nemesis, the isolated self looking out onto the forever separate world. Rather it is the interior world of the chain of transformations surrounding the constant reference at each stage. Computers, and the inscrutable calculations that occur within them, have exactly this kind of interior. We are fortunate that we can peer into this Black Box at any time. Coherence and continuity (provenance) are required and neither is sufficient, for Latour, without the other in describing how Circulating Reference is to work. That resemblance to the original referent(s) is sacrificed at each stage is no great loss. In fact, it is only a major loss if a representational epistemology is asserted. Something of the signal is lost in order to gain mobility, which is why it is important to *maintain a relationship*. One that can carry momentum, much as an electrical signal does. *Work* is carried out. *Translation* occurs between one end of the chain of reference and the other. That it might be useful to have a clearer representation of the original object(s) of study should the chain be broken is not a strong argument against this position. If all is working as it should, "meaning…remains intact through sets of rapid transformations" (Latour, 1999, p. 58).

Our models of correspondence or representation or other similar device are thus not only inadequate to the task at hand, they are also *inaccurate*. They *repeat* the fundamental subject/object division and then act as a means to step across that, rather than accurately seek to preserve information. As Latour says, a fifteen day expedition was all that was required to highlight the muddling that has been carried out by "Kantian philosophy" as well as its forebears and inheritors. This yields a new definition to work with for the single most crucial term:

> *The word "reference" designates the quality of the chain in its entirety,* and no longer *adequatio rei et intellectus.* Truth-value *circulates* here like electricity through a wire, so long as this circuit is not interrupted. (Latour, 1999, p. 69, original emphasis).

In isolation, any one of the artifacts in this chain would be meaningless – such as an isolated piece of code. And the truth of both the chain itself and any individual transformation within it rests on both reversibility and that each stage provides "passage between what precedes it and what follows it" (Latour, 1999, p. 67). Terms such as 'things' and 'words' and the feared gap between them are now, for Latour, "obsolete fixtures of the philosophy of language". Moreover, in contrast to the prior models, these chains have no limit at either end. They are circumscribed, if at all, by the temporary affectation of Black Boxes which can be opened or closed at any time. Previously, world and word were regarded as finite and expressed in terms that require self-enclosure. This is a feature not only actually lacking in the world but also a continuation of the hangover from a need for absolute certainty. In this respect at least, Latour's Circulating Reference definitively *resembles* its object(s) of study more accurately than those of the 'Copernican' traditions. The trail of a 'connection' between a client computer and a server, with information and interaction flowing between them is the very quintessence of Circulating Reference, and to use the latter in order to describe the former means in a quite profound way, to describe through a kind of *direct* representation – where the price paid in return is one of risk and delicacy, requiring continuous motion.

What does the 'work' in the worknet, of providing any kind of intelligible conceptual content to hold the entire networked assemblage together is a *function of distance and connectivity.* And moreover, it is a connectivity that must be actively maintained. It goes beyond a set of propositions, connections and actors that just happen to cohere. It is instead an active *circulatory system*. It is an excellent model of networked socio-technical interactions via machine delegates.

From Circulating Reference to Coherentism

Latour's contribution of Circulating Reference provides a particularly useful epistemological heuristic for grasping, for want of a better term, these 'slippery' mobile references and referents that, routinely, cross divisions such as subject/object. Entities, such as bots and networks, that have a wide range of behaviours, states and, importantly, positions relative to one another lead to a nightmarish situation if one attempts to grasp them using more statically oriented epistemological notions, especially those that focus on crossing the artificial subject/object divide. Such a distinction can of course exist, however within a Latourian metaphysics, it only does so in the same way as every other distinction – it is artificially imposed and requires work to maintain or observe, or both. In many cases it is unnecessarily inimical to the overall epistemological and ontological project – forcibly *purifying* entities that are more realistically observed to behave as *hybrids*, such as the browser-server-webmaster assemblage considered in the case study in Chapter 4.

One particularly useful insight from following Latour's argument here is that many epistemological and ontological problems are created at the level of how their referents are talked about. It should be fairly intuitive to grasp that static conceptions are not particularly up to the task of correctly (or even usefully) characterising the mobile and dynamic. The solution must contain within it the capacity to remain elastic whilst also continuing to capture the meaning and sense of whatever it refers to. This is where Latour's dynamic of work, or cost (in the sense of paying a price, sometimes in terms of swapping properties) provides a significant operational advantage. In short, in order to meaningfully capture and describe our intended objects of study here our methods and concepts really should resemble their objects of study.

To that end I propose using Circulating Reference as an excellent starting point, a heuristic, for outlining our *elastic narratives*. I also propose going *extending* it with the addition of further useful tools such as a version of Coherentism. Specifically, in this case, a variant based on Paul Thagard's 'Computational Coherence'.

Deploying this concept enables us to make use of *elastic narratives* containing within them the ability to stretch, move and yet still usefully circumscribe our objects of study. In practice this means swapping precision and a proof for the crossing of the subject/object divide for mobility. Instead of precision we have a semi-historical account of Propositions – *events* – where some kind of property swap occurs. This would also often entail an accounting of the 'list' of trials of strength our chain of reference has gone through as described above in Latour's view of how something's 'reality' is justified epistemologically. There is no need for a painful epistemological gamble to cross the subject/object divide, because it is crossed continually throughout the chain of reference and by the devices we delegate to measure and record on our behalf. Swapping accuracy for a more general, reliable, veracity it acts as a heuristic enabling the effective description of sociotechnical events, worknets and Black Boxes. Provenance is given and, importantly, *maintained* through the *active* elastic narrative.

However, deploying Circulating Reference, even with its numerous 'helper' concepts (Institutionalisation, Obligatory Passage Point, Property Swapping etc), is an action that is both post-hoc (thus ruling out the possibility of something closer to real time, or even predictive analysis) and, more importantly, *slow*. Whilst it is conceivable that one could anticipate problems beforehand by carrying out this kind of analysis *in situ*, it does not allow for sudden changes in the worknet, especially those that are wrought by reliance on other worknets providing inputs – such as external sources of information. To serially apply this analysis to all of the incoming connections from other worknets leads to a perpetually infinite

regress in any case, so what is required is an additional heuristic approach that can cope with changes and speed and simultaneously allow the assessment of multiple elastic narratives.

Circulating Reference provides a means of dealing with the elasticity inherent in describing ever mobile bots, networks and humans in networked assemblages such as internet connections and electronic trading interactions. What is lacking however is the means to both cope with the speeds at work, how to scale up to larger pictures such as the 2003 Ohio power outage, and also how to distinguish and choose between (or to make consistent) multiple elastic narratives that purport to describe or explain the same phenomenon. Latour asserts that an important part of deploying Circulating Reference is finding ways to foreground or contrast one's intended object of study against a background, or backgrounds, which make it stand out.

Indeed, Latour's specific methodology for applying Actor-Network Theory in the field, as outlined in (Latour, 2005), indicates a clear direction in which to expand Circulating Reference with a viable addition such as Coherentism. Latour states that several things should take place in a good Actor-Network account:

i) the identification of spokespersons (who must constantly produce traces, along with other members of the group for it to even count as one). Groups require effort to maintain and, "Although groups seem to be already fully equipped, ANT sees none existing without a rather large retinue of groupmakers, grouptalkers, and groupholders" (Latour 2005, p. 32).

 ii) Identify 'anti-groups' – part of the effort required for maintaining a group is to contrast it with other groups in order to maintain its boundary.

iii) Assume "that the inquirer is always one reflexive loop behind those they study" (Latour 2005, p. 33). The typical assumption in 'common sense' approaches is that the researcher is *at least* one reflexive loop *ahead* of those observed.

iv) The identified spokespeople should *include* the researcher(s). For their very attention also fixes, and makes the group possible – "any study of any group by any social scientist is part and parcel of what makes the group exist, last, decay, or disappear" (Latour 2005, p. 33). Ultimately, this methodology focuses on attention to context. Action always occurs within context(s) and may be unintelligible without them.

This entire approach is highly congruent with Coherentism, especially with regard to identifying controversies, groups and anti-groups. Computational Coherentism parallels this by identifying multiple collections of cohering and discohering entities and relationships.

Actors giving accounts of their actions will not only divulge postulated forces and other actors driving them, they will also provide novel (if sometimes inconsistent) social theories, epistemologies, ontologies and metaphysics of their own. For Latour, it is absolutely not the job of the researcher to presuppose these, as observed actors will provide them, "[w]e have to resist pretending that actors have only a language while the analyst possesses the *meta-language* in which the first is 'embedded'" (Latour, 2005, p.49). In fact, they can and will provide such a deluge of data that Latour says the social scientist will 'fall asleep' before said actors stop providing it. Circulating Reference can assist in our understanding of how durable or reliable our references are, however it has limited use for dealing with such an information overload, especially if it results in multiple narratives to choose from and evaluate chains of Circulating Reference, most or all of which may appear initially viable.

The threat to the researcher here is that s/he will be overwhelmed with postulated entities (including from their own 'side'). However, Latour argues that controversies *about* agency

and agencies tend to order themselves into obvservable patterns in any one worknet. Thus it is possible, in a realm where there could be an infinite list of groups, to identify the primary patterns of controversy about group formation. Similarly, for the present purposes, it is possible, to apply Latour's idiom, to use "grips" to identify the ways of crediting and discrediting agency in any one worknet. How would one do this? Enter Coherentism.

Coherentism

Competing elastic narratives present something of a problem. Latour's methodology is not up to the task of describing, never mind prescribing, appropriate 'Proposition aware' choices for making informed decisions *between* claims or subjects that are equally steeped in credibly grounded Circulating Reference accounts. Where the latter is successful at all, it remains fundamentally backwards looking in this regard.

Whilst it has been useful to consider how to describe *a* narrative, when numerous narratives take us on different journeys, what we need instead is a *map*. I argue here that the best approach is to use a similarly fast and dynamic methodological concept. How to catch the incredibly fast speeds of computer systems? Why, with another computer system of course. Latour's proposal for Circulating Reference was one that required a change from one end of the chain of reference to another. Mobility was purchased through relative dissimilitude. However, the salient aspects of the instruments used to first capture phenomena, such as the gridding or the pedocomparator were closely aligned with the objects of study in their very properties. A similar alignment naturally occurs in computing and digital epistemologies: deploy a digital instrument to describe or measure a digital worknet.

The earlier discussion of Propositions and Circulating Reference provides a (potentially reliabilist) account of how to trace the provenance of knowledge claims in terms of chains of reference of greater or lesser reliability and endurance. It satisfies demands for the replication

and retracing of any claims made. It also provides a reliabilist *justification* for any claims of veracity made on its basis because of the aforementioned methodology. However, one quickly runs into trouble when multiple accounts are derived this way. Whilst Latour identifies the relative strength of Pasteur's Propositions over Pouchet's for example, the account is highly sensitive to context – specifically at which point in time the debate is considered. The story Latour tells is one of a gradual strengthening of one and a weakening of the other, primarily in the Institutional context. Whilst this is useful in a historical context[31] it is of limited value in anything like a live or near-live context. If there are competing accounts or narratives, it does not matter if the provenance of their propositions can be clearly identified through Circulating Reference. A consistent methodology is required for providing epistemological justification for choosing one account over another and moreover it has to be amenable to capturing and interrogating dynamic and fast sociotechnical worknets. It is a tall order. Classical Coherentism is not up to the task, however a heavily modified version of *computable* Coherentism might be.

As discussed in Chapter 2, one of Bonjour's criticisms of Coherentism was that he thinks such systems could be created in an indefinite variety. This observation is interesting given the purposes of this chapter in resurrecting a form of Coherentism to deal with just such an issue. Multiple coherent systems are envisaged as external constructions (possibly, though not necessarily in the form of clearly articulated chains of Circulating Reference) first before being articulated in a suitable form for internal belief systems.

This is evidenced clearly in the narratives and editorial biases visible in the churnalism data – both mine and that of the Cardiff study discussed in Chapter 5. 'Multiple coherences' are all around us and in constant flux and competition. This therefore isn't an effective objection to

---

[31] And indeed, may be one of the only viable research methods given the visibility of traces in Latour's notion of Institution and the work it carries out in order to preserve one Proposition over another.

Coherentism, but an observation of the world as we find it! Recall the discussion on Latour earlier concerning how we are just plunged into the thick of things and never quite find that perfect starting point. Bonjour (1999) regards the very feature of Coherence as having no obvious connection to any particular notion of truth (beyond, perhaps a vague conception of 'order' being well-disposed towards truth). For our purposes here, this connection was originally sought via Circulating Reference, which *could* have become the standard bearer in any difficult choices between two equally cohering systems. Note however that this is, implicitly, the ontological ground given for supposing some kind of truth conducive component to the overall picture I am presenting and defending here. An adapted Coherentist position is still necessary for providing *justification* in the face of competing evidence from other similarly well-grounded trails of Circulating Reference regardless[32].

Bonjour makes a strong connection between circularity and coherence in order to get his way. Presenting circularity as a kind of self-justifying principle for coherentism that inevitably has no credible justificatory power. The only clear rejoinder to this for Bonjour is the claim that the critic of Coherentism is incorrect in assuming that justificatory relations follow a linear pattern, implying a sort of hierarchical relationship between beliefs. In contrast, it is replied, the coherentist claims that the beliefs "stand in relations of mutual support, but none being epistemically prior to the others" (Bonjour, 1999, p. 123). This could be characterised in masonry for example, where an arch structure has a keystone and surrounding bricks in relationships of counterbalancing force. This, for the coherentist, prevents the relationship being strictly circular.

This counter objection does not satisfy Bonjour however, for he asserts that this simply

---

[32] Related to this latter point is a curious step in Bonjour's argument where he asserts that "probably no one has ever seriously advocated a pure coherence theory of *pure* justification," (Bonjour, 1999, p.122). He contends, rather, that coherentists make an *appeal* to the notion of Coherentism whilst mitigating the primary objections he has identified above. It is motivated by an "essentially negative aim" (Bonjour, 1999, p.123) of avoiding a lapse into foundationalism. Coherentism is thus adopted as the least worst alternative to foundationalism.

moves the circularity elsewhere. I think this is a powerful objection, possibly fatal, *if* one regards coherentism in purely *internalist* terms, as he indeed does:

> In this way, it is alleged, any true circularity is avoided. Such a view amounts to making the system itself the primary unit of justification, with its component beliefs being justified only derivatively, by virtue of their membership in an appropriate sort of system. And the property of the system, in virtue of which it is justified, is of course specified as coherence (Bonjour, 1999, p. 123).

Rehabilitating coherence in a way that can help with apprehending fast moving networked assemblages and associated events (widespread power outages, flash crashes) and in assisting in choosing from the plurality of narratives from trusted sources requires then a characterisation of coherence that is not substantially similar to that conceived by Bonjour. If it isn't a circular, semi-hierarchical relationship then how is the 'mutual support' criterion supposed to be clearly distinguished? Bonjour rejects logical consistency as a necessary requirement here, not least because it depends on so many other processes that are not fully understood and remain controversial, "such as induction, confirmation, probability, explanation" (Bonjour, 1999, p. 123). Humans struggle with these kinds of assessments, especially when speed and complexity increase. Computers on the other hand, when configured and programmed correctly, don't.

Computational philosophy – Pollock

John Pollock has spent more than a decade developing 'artificial intellects' ('artlects') on computers, to test specific epistemological theories. In doing so he also offers a deep critique of contemporary philosophy, particularly modern epistemology. He says, of most epistemological theories that,

They characterize epistemic justification in terms of a general structure and a lot of diverse unrelated principles regarding particular kinds of reasoning (perception, induction, other minds, etc.) without any general account of what ties all these principles together. As such, they are at least inelegant, and cannot help but leave us wondering if there isn't something more to be said which would explain how this conglomeration of principles comes to be the correct conglomeration (Pollock, 1999, p.384).

Pollock also asserts the existence of 'Generic rationality', of which 'human rationality' is but a subset. He considers it an accident that humans are built the way they are and thus only exhibit certain rational behaviours and not others (which were, in turn, evolutionary 'solutions' to 'problems' presented by the environment). Given this, he argues that we can create artificial agents that have arbitrarily chosen rational capacities and put them to the test. The 'arbitrariness' at work here is thus:

Certain engineering problems call for arbitrary choices between solutions that work equally well. Our cognitive architecture probably reflects a great many decisions that would be made arbitrarily by an engineer designing a cognitive agent like us (Pollock, 1999, p. 389).

He argues that, given the tools provided us by digital technologies we need no longer consider that we should mimic human rationality in epistemology. A clear implication from this is that neither should we also seek to mimic relationship structures that are stereotypically human – for with the advent of computing and networking technologies, the speed of relations and changes between them eclipses anything ever seen in a human society previously.

As such, to characterize 'general rationality' (one could also say here, 'general ontology'), we simply state to what design problem it is the solution in any particular context. Human and general rationality may coincide where the 'design choice' made with regard to a particular feature of human reasoning is the only feasible solution; however, we are not limited by such determinism otherwise as philosophers in the past have been - "An anthropocentric bias makes sense if you are doing psychology, but why should philosophy be so constrained?" (Pollock, 1999, p. 390).

The 'armchair philosopher' for Pollock is the philosopher whose primary method of testing epistemological theories is trying to think of counter-examples to the principle under examination. This caricature would classically refer to the radical sceptic and concerned philosophers attempting to rebut them such as Bonjour. Pollock regards this as of limited value in the age of computing. For him, it is possible to precisely specify epistemological or cognitive principles and apply the power of computing to discover counter-examples that would not have been possible 'from the armchair'. Further, it dispenses with two other issues - first Pollock argues that the 'armchair method' suffers from assuming simplicity, in that if a principle has a counter-example, it must be relatively simple and therefore discovered very early on in the reasoning process. Pollock rejects this. If one thing computing has demonstrated in abundance, it is the law of unintended consequences. Even the simplest algorithms can give rise to the most unexpected exceptions when allowed to play out for a long period of time, or to work on novel or unexpected data. Secondly, Pollock argues, artificial entities can engage in bi-directional reasoning. That is, they can reason backwards as well as forwards in a way that would be very difficult for a human beyond the simplest sets of premises, conclusions and inferences - or to put it as Pollock does, the artificial reasoner "reasons forwards from the premises and backwards from the queries" (Pollock, 1999, p. 398).

He goes even further and enjoins philosophers to implement their theories in computable forms if they want to remain relevant for,

> a theory that looks good from your armchair will almost never work the way you expect if you write a program that directly implements your first thoughts on the matter (Pollock, 1999, p. 391).

Difficulties in constructing the 'low-level' (i.e. 'nitty-gritty') principles of one's epistemological theory should lead to modification of the 'higher-level' theory (i.e. foundationalism, coherentism, direct realism etc), if not its outright abandonment. On the longer view, Pollock's work implies that there are likely to be at least several long standing epistemological principles that may be rendered redundant by the advent of computational technology. Thagard foregrounds exactly this kind of approach.

Computational Coherence – Thagard.

Pollock argues a similar case to Floridi regarding the relationship between philosophy and computing: that the advent of computing technology means a fundamental change to how philosophy can and should be done. The 'philosopher in the armchair' is no longer equipped for the computed and networked age – a computer has to be part of the equation now as a measuring instrument as well as an object of study. Thagard adds a further layer to this with the application of computation in psychology (and, indeed, vice versa). For Thagard, more recent understanding of the human mind places inference in the context of a set of cognitive processes that are largely unconscious in nature and operate in a parallel, not serial fashion. This is the context in which Thagard places his work, carrying out a similar project to Pollock but explicitly coherentist in nature, operationalising the work of the philosopher in the armchair for the digital world. As (Pollock, 1999) would argue, computational limits, whilst

having some correspondences, are not synonymous with human limits. And it is the latter that has limited an understanding of events such as the Ohio power outage, or dealing with information overload when having to identify provenance, as with the discovery of churnalism. Tracing a piece of data back to its source on a regular basis would be a difficult task for an unequipped human yet fully possible to automate with computing technology.

Thagard argues that philosophy should be re-acquainted with its lost relationship with empirical enquiry. It should become once again *naturalistic*. Thagard is at pains to distinguish his naturalism from that of (Quine, 1994). He attribute's Quine's approach to the latter's influence from Skinner. Dubbing it a *behaviourist* naturalism rather than Thagard's *cognitive* naturalism. He argues that it supersedes both analytic philosophy and phenomenology whilst embracing a coevolution of philosophy and the relevant sciences. Interestingly, Thagard explicitly identifies common 'naturalistic epistemology' ground with Latour and Woolgar. He calls their empirical Actor-Network Theory emphasis on mapping social associations and relationships 'social naturalism'.

Thagard does see some anticipation of his own work in Western philosophy, though believes that it was not possible for many philosophers to properly articulate without digital computing technology. This factor making many of Bonjour's criticisms have much greater force before the computing era. Whilst Thagard argues that philosophers from Hegel onwards might have entertained some kind of coherentist view of justification for example, he points out that such theories could not be properly realisable until the modern age when they became computable. Unfortunately, simple serial inference in logic and foundationalism in epistemology still largely dominate the analytic tradition. Whilst a serial logic is able to model, and to assist, humans engaging in specific deliberative reasoning, it does not capture the nature of our cognition at all. What should be a tool is in fact a dogma.

So what precisely does computing technology bring to the table that would both surpass the kind of criticisms levelled at Coherentism by Bonjour and also enable the selection of one narrative over another when provenance is in doubt? Thagard argues we can computationally maximise coherence through "maximising satisfaction of a set of positive and negative constraints." This requires determining characteristics that would make two 'elements' either cohere (fit together) or incohere (resist fitting together). He lists examples of these relationships as follows:

> Coherence relations include explanation, deduction, facilitation, association.... Incoherence relations include inconsistency, incompatibility, and negative association (Thagard, 2000, p. 17).

To decide a coherence problem, (e.g. the choice between the Daily Telegraph's version of events and the Daily Mirror's), the solution is found by dividing elements into two sets 'accepted' and 'rejected'. Coherence is maximised through the maximal satisfaction of both types of constraints. This is an important point. Coherence is still maximised even if the most constraints that are maximised are of the incoherent (rejected) set. This occurs as extraneous material is removed from the cohered set, however small it is. There is also the further option of carrying out the same procedure on the rejected set, perhaps with different hypotheses, to see if that set has some kind of coherence, separate from one's initial theories and preconceptions. One could thus end up with several coherence sets (though this is not something that Thagard explicitly pursues). As a crude measure of coherence, Thagard proposes that W is the total weight of all constraints satisfied following the two coherence conditions. Maximal coherence has been reached if there is no other division of the various elements into A and R that would result in a higher value for W (bearing in mind that in some runs, some constraints may *not* be satisfied).

Unfortunately to carry out such complex like for like comparisons on even a relatively simple system is very computationally expensive. Even with computing power continuing to follow Moore's law, and the development of novel Big Data techniques such as MapReduce, the comparison of thousands of individual elements to one another, using various coherence relationships as would be required for assessment of even a simple human belief-system, is so computationally expensive as to be inappropriate for a 1 to 1 comparison for every element and every possible state. This is the kind of intuition Bonjour was following in dismissing the possibility of a conscious apprehension of all held beliefs – at least for a human. The computational expense comes from insisting on continuing decision making and pattern matching in a linear way.

Following the traditional sequential means of philosophical argumentation would make pursuing a coherence strategy, even a computed one, as close to impossible to carry out in good time as to make abandoning the project inevitble. Thagard puts it into figures in this way:

> for n elements there are 2n possible acceptance sets. A small coherence problem involving only 100 propositions would require considering 2^{100} or 1,267,650,600,228,229,401,496,703,205,376 different solutions (Thagard, 2000, p. 26-7).

He argues that nature however, has inspired a potential answer in the form of neural net models. Whilst it is still not appropriate to compare software-based neural nets directly to biological neural systems, it might be appropriate to compare them to models of conscious human cognition. There is also substantial mileage in using them to enhance actor-network accounts, as the worknets traced by ANT practitioners surely follow similar patterns and behaviours in real time. Weights can be given to particular links in such a net that resemble

the kind of pattern matching a human would do, and also take account of particular philosophical and psychological views of cognition. Importantly, for cases of dealing with biased (or otherwise corrupted) information, this provides a safety valve as even if the specific form of the bias or corruption in the information is unknown, it can be caught by both looking for incoherence and applying a purposefully discohering weight to the source in question. One could potentially also apply this to one's own biases if one has sufficient self-awareness to be able to articulate what they are. It also enables an elegant means of including definitely trusted inputs into a coherence model.

For example, most philosophers and psychologists would want sense-data to be privileged. In a neural net this can be achieved by weighting those relationships or pathways accordingly, so sense-data is more likely to generate coherence, whilst at the same time holding out the possibility of it being rejected if the remainder of the system coheres against it (thus answering Bonjour directly). This is as it would (or perhaps *should*) be in the example of a dream or hallucination. The algorithms deployed for such nets follow a *connectionist* philosophy, where updates are simultaneous across the system as opposed to a serial update where every possible system with a single change is compared to another, leading to the enormous number of calculations identified above.

To be clear as to what connectionism refers to in this context - a connectionist approach uses neural network algorithms. Using this method, every element has a corresponding node, a 'unit' that is a node in a network of units. Such a network is supposed to crudely represent a network of neurons. For every constraint between elements, an excitatory or inhibitory link is established in the corresponding units. These links subsequently determine the level of 'activation' of each unit. An ingenious addition is a special unit that can be used to fix excitatory links that belong to some special class of units. For example, in some models, those coming from sense perception or empirical evidence, thereby giving them greater

weight (higher activation in this context) as their influence propagates throughout the network.

All of the units are then updated in parallel with the activatory and inhibitory impulses spreading throughout the network units (nodes/elements). It is important to note that whilst all the links / constraints are symmetrical between elements / units, their pattern of activation is not because of the influence of the special unit. The activation updates continue until the network has 'settled'. That is to say that the activation values of the numerous nodes / units no longer changes. If the final value for each unit is above 0 then it is regarded as accepted, and below 0 rejected.

This kind of model has tremendous intuitive appeal and can be shown to model real-world reasoning processes, at least where it is possible to bring them to consciousness and articulate them. Proponents of neural networks, including Thagard, also argue that it models such unconscious processes too. Its main drawback is that it is not a guaranteed method of maximising coherence, and on this count may underperform approaches such as semidefinite programming (and of course the computationally prohibitive exhaustive approach)[33]. There is also no guarantee that the network will settle at all or that it will be computationally affordable, though generally in practice it will. However, despite the dearth of mathematical guarantees, it does provide significant insight and explanatory performance.

Such an approach also makes the traditional opposition between Foundationalism and

---

[33] These other approaches are described as follows: *Exhaustive search algorithm -*
This algorithm is essentially the nightmare scenario that animates concerns about computational limits. It is time consuming because it is comprehensive and will yield the most accurate and complete result. It would consider *every possible combination* of rejecting and accepting elements, resulting quickly in out of control 'solutions' that cause more issues than the problems computational coherentist approaches are attempting to solve in the first place.
Semidefinite Programming (SDP) -
This is a complex mathematical area and approach in its own right. It begins as a similar approach to the incremental algorithm, applying a linear function, but optimises it against matrices. Again Thagard reports some success in using such an approach, especially as it can guarantee a minimum level of the optimum W, however it does not appear to usefully model real world reasoning.

Coherentism somewhat redundant and in need of rethinking. For it occupies the space of both On the one hand the coherentist view that there are no indubitable truths and that justification is provided by the extent to which beliefs (or other 'elements') fit with one another. On the other it encompasses the foundationalist view that not all beliefs (or 'elements') are equal in terms of contributing to justification. This is straightforwardly included by the ability to weight and privilege connections.

Maximising Coherence

Coherence, for Thagard, is maximised through the process of 'maximal constraint satisfaction'. What is 'maximal constraint satisfaction' in a Coherentist context and how can it help us? As mentioned above, constraints are a particular type of relationship between two or more elements. They are divided into positive and negative constraints. To maximise coherence, as many of *both* types of constraints have to be satisfied as possible. They are dynamically interacting. The constraints are characterised via the elements that make up the model. These elements could be representations, propositions[34], concepts, goals, image parts, actions and many other things besides. If they fit together, they cohere. Constraint relationships that indicate coherence, or fit together (positive) could include *explanation*, *entailment*, *deduction*, *facilitation* (e.g. action *a* facilitates goal *b*), *association* and other similar types. By contrast, constraint relationships that indicate incoherence or not fitting together (negative) could include *inconsistency*, *incompatibility*, *negative associations*, *competition* and so on. The specifics depend on the type of elements being considered and the type of coherence being applied.

Following this, the elements are subsequently divided into those that are accepted (relatively cohere) and those that are rejected (relatively, do not cohere). A positive constraint is

---

[34] Which could include 'small p' propositions as are commonly understood (i.e. statements of affairs or premises) or Whiteheadian Propositions (events involving the meeting and interaction of two or more worknets).

satisfied on this basis if both elements are accepted or both are rejected (remember, this is *coherence* we are talking about here – if *a* then *b*, *not a* entails *not b*). On the other hand, a negative constraint is satisfied *only* if one element is accepted and the other rejected. Maximal coherence is reached when elements are divided into 'accepted' and 'rejected' groups in such a way that the maximal amount of constraints have been satisfied. Two different media narratives of the same event for example, may include and exclude different details. Modelling which is the most coherent and which is the most incoherent narrative could be quite informative. As could subsequently altering the values for privileged activation / inhibition representing assertions of differing levels of trust in one source compared to another.

Using Thagard's methodologies and algorithms, all such constraints are apprehended and assessed together, though this is a maximising goal and as such not all constraints may be able to be assessed simultaneously. This means that, especially in larger representations of element-constraint relationships, local relationships of coherence and incoherence are taken account of. So for example if *a* explains *b* then it is desirable that if *a* is accepted, then so is *b*. The resolution of the latter's constraint satisfaction then radiates out to the remainder of the model.

This is expressed in Thagard's two primary coherence conditions:

> If (e_i, e_j) is in $C+$, then e_i is in A if and only if e_j is in A.

> If (e_i, e_j) is in $C-$, then e_i is in A if and only if e_j is in R

> (Thagard, 2000, p.18).

'A' represents the accepted set, 'R' the rejected. Elements with some kind of constraint relationship are represented with e_i, e_j and $C+$ / $C-$ represent positive and negative constraints

respectively.

In addition, to add some crucial fine tuning to the model, a further factor 'w' is added. This represents the weight, (the strength), of the constraint. And in mathematical terms this is reflected in Thagard's constraint relationships as a real number between -1 and 1. Some elements may more strongly or weakly cohere and this weight enables reflection of that. It is important to point out here that a relationship of coherence means that the constraint is *symmetrical* – i.e. there is no order of precedence to assessing the constraint between two elements (as may be assumed for example in the circular model above in Fig.11).

The response to the above concern regarding how to establish the elements and constraints for whatever kind of coherence or 'truth' one is seeking to model may have already occurred to the reader as it has already been covered: Prior to applying computational approaches *pace* Pollock and maximising coherence *pace* Thagard, we establish the elastic narrative(s) through Circulating Reference and Propositions. The components of an account expressed in terms of Circulating Reference naturally provide the elements and constraints – our worknets (or Black Boxes) and relationships/associations, respectively.

Both Thagard's overall approach to Computational Coherence and the various types of justification sought (with corresponding constraints) enable significant amounts of flexibility in shaping the tool to the task in hand. The specific method chosen is selected *because* it bears epistemological and ontological similarity to its object of study and thereby becomes a much more accurate measure in its own right[35].

---

[35] Unfortunately, it is out of the scope of this thesis to consider each of Thagard's models in turn. The subject matter here focuses on methods of description and apprehending rather than models of reasoning, though there is of course plenty of natural crossover. For posterity though it is worth noting that of the above kinds of coherence problem, for the purposes of this thesis, we would most likely be concerned with those types that impact explicitly on epistemological goals and issues (and, in addition, those for which Thagard and others have developed workable algorithms to assess). Thagard lists these as "explanatory, analogical, deductive, perceptual and conceptual". Whilst all five are of interest, the first three would be the most relevant to this inquiry.

Summary

When discussing the artefacts and symbols used as representations in a chain of Circulating Reference, it is crucial to realise that these representations are themselves combined with others, forming new relationships, summaries and manipulations that in turn form higher-level and more abstract representations. Each level of abstraction makes the representations more general and takes them further from their original objects of study. An ideal world for Latour would be one where there were no leaps between data and theory nor theory and application, only a set of discrete and minute steps. Latour asserts that there should be no action at a distance contained within our theories, descriptions or explanations (Latour, 2005). However, most humans and researchers operate *as if* there is. And because the epistemologies and ontologies present in software are determined by their human creators, many programmes also operate with the same built in assumption. Latour draws attention to this and asks one to return to the minutiae and not make these leaps – to observe first and ideally not interpret, or at least ensure interpretation comes second to observation and is clearly demarcated from it. This is the very core of both Latour's empirical method and the central battleground for his more explicitly philosophical concepts, particularly in ontology and epistemology.

Circulating Reference provides a consistent way to describe, trace and even measure these movements and is especially appealing for applying to digital and networked assemblage contexts. The example used earlier of an ongoing internet connection is particularly apropos. It is in constant flux, and yet *something* endures – the chain of relationships. The 'reality' of the connection is further strengthened, or weakened, by the routes and resources it has

available (which would be 'allies' in Latour's terminology). The characterisation at work has an essentially elastic nature.

In order to extend such accounts for worknets where the speeds at work make it difficult to assess how chains of reference are being constructed, or where it is difficult to choose between competing narratives regarding their construction and provenance, Computational Coherentism provides a strong adjunct for both decision making and epistemological assessment.

I now turn to consideration of several empirical examples in networked assemblages where such problems beset the investigator and discuss how the insights and approaches covered so far may help unpack, explain and articulate them. The first of these case studies is an interesting meeting point where human internet users, bots, web servers and their human support staff all converge – the 'robots.txt' file.

Chapter 4: Case study: Robots.txt

Non-human agents, 'bots' currently make up the majority of web traffic activity (Incapsula, 2014). Many complex examples of networked assemblages or events such as search engines and electronic market-making respectively occur with them at their centre. It is important to be able to consider the 'bot' outside of its immediate source code and to visualise it in broader terms of the environment in which it is placed. How the underlying technologies and individual instantiations of them are conventionally Black Boxed (subsequently leading to ethical and epistemological difficulties). One way to apprehend networked assemblages that include bots by observing them in these ways as they engage in Trials of Strength, something that informs the empirical basis of this chapter.

The primary object of study in this chapter is the 'robots.txt' file. Something that provides an Obligatory Passage Point for any networked assemblage where websites are a key feature. Any web spider seeking to index or query a website is expected to consult this file. Whilst it is an informal standard, it is widely, and voluntarily, adopted. Also referred to as an 'exclusion standard'[36], it gives instructions to any bot browsing it regarding what is and is not permissible to scan and will often include other rules such as the amount of time that should elapse between each query to the site. It may even blacklist particular bots (for example the Google Spider[37]) for some possibly arbitrary reason.

Michael Schrenk in *Webbots, Spiders and Screenscrapers* (Schrenk, 2012), describes a big moment of discovery for him decades ago when he first realised that the internet is accessible

---

[36] Contrasted with an 'inclusion standard' such as the sitemaps protocol that provides bots with a ready-made map of the site, thus minimising time spent requesting pages and giving a clear structure to follow. The 'sitemaps' protocol is specified here: 'What are Sitemaps?' - http://www.sitemaps.org/ (last accessed 22/09/13).
[37] See Google's description here: https://support.google.com/webmasters/answer/182072 (last accessed 30/09/2014). This is the primary workhorse for Google's ability to index sites and their continually changing context. It is noteworthy when a robots.txt directive specifies to block this (or any other major search engine) bot.

without a browser. On the fly programs and automated agents can be used in the stead of a browser, and in his book Schrenk offers instructions on how to make and deploy them. Describing what is going on in these complex socio-technical systems is still difficult however, even if the technical methodology for making and deploying bots is relatively straightforward.

How bots such as those described in Chapter 2's taxonomy interact with each other, human users and computer networks is an ongoing ethical flashpoint and legal grey area. It is also an area amenable to modest empirical analysis. Description of such networked assemblages ought to not only include the technical and socio-technical systems but also tease out the relationships with the humans involved, as their understanding and view of 'bots' will have an effect on the behaviour and thus impact of the 'bot' in question.

Bots of the various types listed in the taxonomy were central participants in the 2010 Flash Crash and the Google News incident. Many of them would be following complex decision-making algorithms and their inner workings may remain a mystery as the strategies and models behind them will be difficult to access. On an individual level, they are only viewable from the outside as Black Boxes, though a significant effort has been made by Rishi Narang in his *Inside the Black Box* (Narang, 2013) to describe the strategic operation of High Frequency Trades (HFTs) in general detail. Narang's intent however is mainly to head off what he perceives as unjust criticism of, and hostility towards, HFTs made on the basis of ignorance. He provides an in-depth description of the kinds of activities one *might* use HFTs for. However, this does not shine a light into the general mystery of how we are to grasp the dynamism at work between so many human and non-human entities. Even had we the source code available, it would be of limited use because the intentions behind a bot's algorithm and predicted range of outcomes frequently don't survive contact with other bots, networks and

indeed, humans. Of more use is an approach that can be used to sketch the boundaries and key features of these kinds of interactions in general terms.

The bots that are conventionally used to search, index and query websites are highly amenable to such analysis. They are easy to construct, operate in a publicly viewable domain, operate at speeds that are more comprehensible to humans (in seconds rather than microseconds like HFTs) and exist in a sociotechnical ecosystem where both humans and bots are in play on a daily basis. We can use these to shine a light into ('whiten') the Black Boxes at work in the much more volatile domain of the Flash Crash.

Choosing Bots

Following something of a tradition in science and research, the acquisition of the primary source material for this chapter's topic was somewhat accidental, even serendipitous. Originally, I planned to spend significant time on studying the behaviour and interactions of other parties with numerous simple Spiders, Scrapers and Mechanizers that were custom built for the purposes of automating press release 'churnalism' detection (see Chapter 5). Studying the activities, and interactions with, these bots enables the foregrounding of the multiple bot Black Boxes that conventionally work invisibly.

As mentioned in Chapter 1, two things occurred part way through the research that directly interfered with this plan. Firstly, the third party tool and online resource – the 'churnalism engine'[38] - that was to be leveraged by the aforementioned bots, completely changed its method of functioning (in fact, neatly reversing it). Secondly, one of the larger batches of target websites was a list of major government departments. They went through a major restructure, pulling all of the content from what were originally diverse websites into a single site. Ironically both events demonstrated the very real problem at the heart of this inquiry –

---

[38] www.churnalism.com

that the worknets under observation are delicate and subject to (sometimes arbitrary) change at a moment's notice.

Serendipty interjected at this point. Indeed, given the object of study's mobile and dynamic nature, it seems that serendipity itself necessarily becomes part of the research method. A useful resource was already available – the gathered 'robots.txt' directive files. At least one of each had been collected[39] for each of the target websites and was required as part of the process in creating correctly constructed bots for each site. Moreover, the literature on using these files as an object of study is limited and primarily concentrated in Computer Science rather than Software Studies. Large scale surveys of robots.txt files have taken place, for example (Sun, Zhuang and Giles 2007; Kolay et al. 2008). However, they focused primarily on issues such as bias towards particular search engines and counting the number of sites that used the file. Interesting to be sure, however such studies were not particularly informative for exploring the sociotechnical angles. Nevertheless, as empirical descriptors of what expectations were placed on incoming bots by website creators and maintainers these files yielded an interesting and understudied area of research in their own right. They therefore became the primary focus of this chapter.

Target websites

The selection of websites for scraping was originally inspired by the 'churnalism' research of (Lewis, Williams and Franklin, 2006).They manually traced press release content that appeared in broadcast and print media back to its source (see Chapter 5 for more details). The prospect of automating such an analysis became realistic when the 'churnalism' site was launched in 2011. A second reason for looking at both at the issue of 'churnalism' and the particular selection of sites was the fact that the worknets in play during the 2010 Flash Crash

---

[39] As discussed further down, even these modest files are subject to sudden, sometimes highly revealing, changes. In a couple of cases this change over time was captured.

and the Google News incident. Some of these sources of news were also automated and designed explicitly for machine consumption (Steiner, 2013). Incorrect or misleading information feed through these channels could obviously be critically important to what happens in the socio-technical domain of electronic trading as both human and machine decision making is based upon it.

A triple purpose was thus served here. First, bots were constructed to automate a potentially useful function that had previously been done primarily manually (tracing the provenance of information to/from press releases) – an exercise that yielded some interesting results in its own right (see Chapter 5). Secondly, the research focused on a domain that was directly applicable to the complex networked assemblages of concern here – the provenance of information that could alter decision making for algorithms. Thirdly and most importantly, it provided a neat example of the kind of socio-technical interactions, albeit somewhat simplified and substantially slowed down, that were at work during sociotechnical imbroglios such as the 2003 Ohio power outage and the 2010 Flash Crash.

Lewis et al. originally categorised their work into 'subject focus' areas crime, domestic issues (NHS, education, immigration etc), politics, business/consumer news, health/natural world, entertainment/sport, accidents/disasters, defence/foreign policy and 'other'. Owing to their methodology, they were starting from the stories they selected in the news over a two week period and tracing them back to press releases. Categorisation was thus led by the topic, usually easily identifiable, of each story. By contrast, the 'churnalism engine' was *originally* designed to accept press release copy and compare it to the output of the majority of UK media already stored in the 'journalisted' database. For categorisation I therefore had to classify organisations that were issuing press releases. I wanted to capture as much of the categories covered by Lewis et al as possible whilst also cleaving to the kinds of topics that would feed directly into electronic trading algorithms.

Originally the categories were to encompass 'business', 'government', 'health/medicine', 'environment/disasters' and 'media'. The latter category was added as media outlets would often recycle each other's stories (Davies, 2009); covering their output separately would potentially enable the provenance of stories to be properly refined. In practice, only a handful of robots.txt files were kept from three of those categories, as curiosities, because they had interesting features – 'health/medicine', 'environment/disasters' and 'government'. The first category suffered because press release outputs seemed to be dominated in the UK by the BMA (British Medical Association) and little else. The second because press releases were dominated by activist groups and the third because the UK government retracted all press release activity into a single site. The two categories that remained relatively intact for worthwhile (i.e. comparative) analysis were 'business' and 'media'. 'Business' sites were selected according to the top ten companies in the FTSE index, and 'media' on top ten circulation figures. Several had to be discarded from the 'Business' category as they had either no discernable press release output or robots.txt file. British Chambers of Commerce and CBI were added instead to this category as they produced a significant amount of press releases. This was fortunate in one respect. These two categories contain the news inputs that most strongly influence trading and energy distribution decisions (by humans and machines).

The results from studying the numerous robots.txt files can be found in Appendix 2. The primary features identified are as follows: 'Crawl Delay', if given, is the amount of time a bot is expected to pause between requests. 'Restricted folders' indicates if bots are forbidden from any folders (the particular details of this in some cases are more interesting than whether this is simply a 'yes' or 'no'). 'Banned agents' specifies whether the file forbids any *specific* bots from a particular activity. And finally 'Sitemap' identifies whether a link to a formal sitemap (the 'inclusion' standard) is provided.

A directive to restrict use of the search facilities on websites, or variants of it, is the single most common instruction across all of the robots.txt files collected. On the surface this may make intuitive sense – search facilities can be quite demanding on the server as they usually require dynamic access to a back end database. Machines can make thousands of requests a second, imitating the sending of a completed (clicked, selected etc) form. There is thus a *prima facie* case for restricting bot access in the interests of preserving bandwidth and server time. However, when considering the different uses to which such a bot may be put, it is not quite so straightforward, especially if the needs or purpose of the site changes.

One of the more interesting elements in researching the role of robots.txt files is not just the differences between site policies exposed, it is also the fact that they can change over time and in response to external events. A famous example is that of the whitehouse.gov website robots file – which changed when the Obama administration originally took over from the Bush administration. Cory Doctorow identified the changes in his *BoingBoing* blog (Doctorow, 2009). Overnight, the file changed from 2400 lines of instructions to just two. What made this of especial interest was the fact that the instructions to web spiders forbade the indexing of many topical pages, including on Iraq and 9/11 related information. This would have meant in practice that all of the major search engines, given that they deploy 'well behaved' bots that respect the instructions in the robots.txt directive file, would not have indexed the pages and they would not have turned up in searches as a result. In fact, the only way they would have appeared in search results would have been if they had been referenced by a third party site or article, and even then the coverage would have been sporadic.

Yesterday, the robots.txt file for whitehouse.gov had ca. *2400 lines* worth of files and directories that search engines were not allowed to index. Today, the file is two lines long: "User-agent: *" and "Disallow: /includes/"

```
User-agent: *
Disallow: /cgi-bin
Disallow: /search
Disallow: /query.html
Disallow: /omb/search
Disallow: /omb/query.html
Disallow: /expectmore/search
Disallow: /expectmore/query.html
Disallow: /results/search
Disallow: /results/query.html
Disallow: /earmarks/search
Disallow: /earmarks/query.html
Disallow: /help
Disallow: /360pics/text
Disallow: /911/911day/text
Disallow: /911/heroes/text
```

FIGURE 1

SNAPSHOT OF THE WHITEHOUSE ROBOTS.TXT FILE (DOCTOROW,2009).

Interestingly the file has changed again since then – the latest version (in Appendix 2), is now sixty lines long, though only two thirds of those lines are instructions to spiders (the remainder being comments or formatting). The change between the previous two versions was obviously political, however the latest version looks like it has been changed automatically as a result of a Drupal site upgrade (Drupal is the underlying content management system the Whitehouse site runs on)[40].

---

[40] The Whitehouse webmaster may not even be aware the robots.txt file has changed, especially as additional restrictions are only added to sensibly protect core Drupal directories from unnecessary indexing (a structure which is likely to be identical from one Drupal installation to another and therefore is a complete waste of bandwidth and computer cycles for all parties if it is indexed).

There is a similar change visible for the DEFRA and Environment Agency robots files (see Appendix 2). As mentioned above, in the midst of this research the DEFRA website was transferred to a new domain – gov.uk, joining all of the other ministerial departments in the UK government under a single website, with each department having its own subdomain. The old robots file forbade, in a similar fashion to the new Whitehouse file, indexing of core software directories. The new file is a generic robots.txt file designed to cover the entirety of the gov.uk domain and as such has different priorities now – primarily stipulating restrictions on bots from using the search facilities (of which there appear to be several on gov.uk), which will be primarily *mechanizers* (normal spiders won't access a search page beyond the links and settings already present by default).

Always restricting automated use of the search facility as a default, however, could be problematic. Take the site jobs.ac.uk for example. Its robots.txt file forbids automated use of its search facility. The rationale is that competing jobs websites could easily use their bots to search for and copy job adverts and benefit financially from doing so (charging fees to users, advertisers etc). However, what if the bot author was looking for a job and wanted to optimise the finding of suitable positions? Jobs.ac.uk offers RSS feeds, but these are broken down on an academic domain by domain basis and offer no immediate means at present to select other options such as location. Whilst the RSS feeds could be scraped and subsequently processed by a dedicated bot, it is much less effort to write a single *mechanizer* that carries out the requested searches, with all the desired fields, and across domains, dumping new results that match all of the criteria into a text file or database. This may even save bandwidth on both ends. Yet it would, technically, be a badly behaved bot, operating in defiance of the robots.txt instructions. This is in spite of the fact that such an action actually optimises the *intended function* of the site – to efficiently match candidates with vacancies. That is to say, there are no victims in this scenario, only winners. Such a case highlights the fact that the

intentions of the first movers (or 'provocateurs') in any such complex assemblages (bot, bot user, server, webmaster) can matter as much as the raw actions carried out by the bot (or webmaster).

One of the most interesting findings where searching exclusions are concerned is the Financial Times file – it provides comprehensive restrictions on Googlebot, but no other spider. There aren't even rules specified for all other bots (designated with 'User-agent: *'), which is unusual and raises questions regarding why Googlebot is singled out. On that single criteria, this robots file is certainly an outlier in the available dataset. There are other similarly interesting examples though. Both the Daily Mail and Environment Agency files have a long list of forbidden articles for indexing. This is also unusual and possibly worthy of further investigation regarding the reasons for what is effectively passive censorship as per the Whitehouse robots file discussed above. Ordinarily entire directories are designated as uncrawlable rather than individual articles. This is particularly odd for a major media organisation given that it is usually part of their *modus operandi* to ensure their stories are circulated as widely as possible, especially as many now have to experiment with new models of delivery that incorporate maximising clicks on their site[41]. In an important respect here the chain of Circulating Reference is broken. The articles in question may be cited elsewhere on the internet by authors who read them when they were originally published. However, using search engines to find them could be problematic as the robots.txt file forbids their indexing.

Surprisingly, the specification for delaying bot queries ('Crawl-delay:') is only present in a handful of the files. This is possibly ill-advised because a bot writer is much more likely to comply with a delay request than a directive that forbids accessing certain directories or files.

---

[41] That UK media outlets would go in this direction is a development anticipated by (Lewis, Williams and Franklin, 2006).

As a result, if the primary goal is to preserve bandwidth or forestall anything that could be accidentally regarded as hostile behaviour it is a better choice to specify a reasonable delay than simply guess which directories will attract the most traffic. Note the 'reasonable' judgement– the variance was quite interesting amongst the files. The gov.uk site specifies a 0.5 second delay which is quite reasonable – a bot could still carry out tedious work at a much faster speed than a human at this rate. On the other end of the scale, the British Chambers of Commerce site requested what can only be described as a ridiculous 90 second delay. Obviously this is much slower than a human browsing the files. Perhaps the webmaster benchmarked this on the basis of how fast a human may read a typical article on the site. What is very likely however is that any custom written screenscraper will ignore that delay, making it effectively redundant. The others with a specific delay requested 10 seconds (IES and Greenpeace UK). This is still a fairly significant delay though one which more custom bots are likely to comply with.

The advice in the literature on this issue (what little there is) remains divided. For example, for a recommended crawl delay, (Schrenk, 2007) recommends three seconds, whilst (Rhodes and Goerzen, 2010) recommend ten. No rationale for these specific timings is given, though how and why to treat target websites with care in this respect is explained.

Several of the target websites had no robots.txt file at all, with the highest concentration amongst the 'business' category - three of the ten largest companies currently listed on the FTSE selected lacking the file. All three (HSBC, Royal Dutch Shell and Diageo) had the sitemap inclusion standard, however this is hardly significant as the vast majority of customer facing websites will be built on an established platform or template that will include the sitemap as standard. There are two strands of reasoning that may explain this:

i) The companies are not short of resources and as such are not concerned about restricting the bandwidth or computer cycles that may be used up by a 'greedy' *spider*. Similarly, corporate security systems are likely to be sophisticated enough to immediately detect hostile access or form requests. Smaller organisations have to be much more reliant on bots complying with the robots.txt file.

ii) Much larger budgets also mean a great deal of attention can be given over to communications – what is published on the website may go through a much stricter quality control process than elsewhere and also pushed out via well organised social media and media contacts policies that mean fewer people will be motivated to write custom bots in order to find the information they require (Rigby, 2008). It is a fair assumption that even without directives from a robots.txt file, the dominant search engine spiders will be relatively well behaved regardless.

These files make for an interesting analysis as the de facto object for mapping out the potential interactions of humans and bots in socio-technical networks, with the website as the primary point of interaction. The majority of human centred (initiated) internet activity is carried out via the intermediary of the browser. Whilst generally a Black Boxed activity for day to day interactions, automated interactions on the web are nevertheless Black Boxes that are easily opened or 'whitened'[42]. And if attempting to trace a chain of Circulating Reference, whilst any break in the chain is generally catastrophic (normally stopping the communication or process entirely), it is also usually straightforward to identify and rectify. Bots can be, and are, routinely deployed to carry out actions on both internet and internal networks. Moreover, they often serve to mask the number of intermediaries involved; intermediaries that are effectively invisible to the casual user of a browser. Indeed, one of the primary motivations

---

[42] Even activity taking place on an internal network - whether through use of a corporate intranet or using a file explorer (e.g. windows explorer) to access networked resources, generally takes place through a similar kind of browsing intermediary – at least for humans.

for programmers deploying bots in such contexts *is* to sidestep intermediaries, one of which can often be the 'robots.txt' directives file.

Bots are tools, sometimes effective intermediaries themselves, that in the traditions of lower level programming practices bring the programmer or user closer to the machine. Take the example of the robots.txt files for The Sun and The Times. One of the forbidden bots, or rather, bot *domains*, is anything launched from newsnow.co.uk. The latter is a useful site for users to aggregate stories from multiple (worldwide) news sources according to whatever topics they are interested in. The site uses bots to scrape the headline content from dozens of sources. The bots thereby provide a massive labour saving intermediary force, albeit one that cannot provide its users with content from The Sun or The Times. The site draws on so many sources – including local newspapers and dedicated subject sites (e.g. Techcrunch) that one may not even notice the absence of News International publications. It also means that information from these publications will not reach many of the more complex bots that seek to assess sentiments, market information and so on. Bizarrely both (almost indistinguishable from one another) files have specific instructions regarding the use of bots on the site that are clearly meant for humans:

```
#Spidering is not allowed by our terms and conditions
#Authorised spidering is subject to permission
#For authorisation please contact us - see
#http://www.nisyndication.com/about_us.html
```

FIGURE 2

COMMENTED SECTION IN THE SUN ROBOTS.TXT FILE CLEARLY MEANT FOR HUMANS (SEE APPENDIX 2)

A bot will ignore the above commented out section, and although it is recommended - e.g. by (Schrenk, 2012), many bot programmers will not even look at the file. Both of the News International files considered contain these identical restrictions. The aim is clearly to attempt

to manage the automated flow of data regarding published articles beyond the publication's websites, channelling it ideally through their syndication programme in the link above. In spite of such restrictions, there would be nothing to stop a competent bot writer from scraping content (and doing so whilst masked as a browser using human, not a bot). The rate of publication on either site occurs at a modest pace and as such could easily be catalogued (and potentially shared) by a human with the time, or a bot pretending to be a human. Such flashpoints of interaction create many potential (and actual) ethical and legal imbroglios. They also clearly frustrate the usual chain of Circulating Reference for linking news items and facts back to their source, especially when assisted or done entirely through automated means.

More considerations on 'robots.txt'

The rules set out by robots.txt files are entirely informal and the only method of censure for breaking them (at present) is for the IP address from where the offending bot(s) are launched to either be blocked or have limited content served to it. In extreme cases the owners of the recipient server could pursue legal action via legislation such as the Computer Misuse Act (1990, amended 2006) if it could be demonstrated that a particular bot or bots were actively being used to disrupt the normal activity of the server[43]. As a domain of behaviour and normative concerns it is extremely messy and as such provides a very useful empirical basis for illustrating the task of articulating a general account of fast moving networked assemblages.

---

[43] An excellent example of this was the BBC program, 'Click' inadvertently falling foul of the CMA, as detailed in (Carolina, 2009), *Why BBC Click Violated the Computer Misuse Act*.

One important potential blurred line here is that browsers can be modified with bot-like additions that enhance the capabilities of the user[44]. These may enable lightning fast interactions with websites at the behest of the user. Such software already disrupts the intention behind the robots.txt file to regulate the load on the website. It also raises questions regarding identity. According to the informal rules of conduct a bot is supposed to identify itself clearly (and if possible give a clue as to its intention – e.g. 'link scraper'). This is to assist the target site's webmaster in distinguishing bots from humans on the assumption that humans should not only have priority in accessing the site, (most sites are oriented this way, at least visually, towards humans though by no means all), but also should generally provide a lighter load on the website server. This is clearly of limited use given that not only can a bot easily imitate a browser, but a genuine browser's activity can switch from the usual plodding pace of a human reading and clicking links in an instant to a blindingly fast search and pillage of the site's media if a download manager add-on is activated. This makes planning provision on a website even more fraught as the speed of access (and bandwidth demands) can change in an instant and it does not matter whether, following some particular criteria or other (e.g. initial request frequencies, form filling behaviour) whether new visitors to the site are initially correctly categorised as 'bot' or 'human'[45].

As a concession to bot-writers who need automated access to a site's data, many sites offer an API to use whilst disallowing any bot parsing of their pages. An API (Application

---

[44] A particularly popular add-on across different browsers for example is a download manager – prior to more sophisticated browser software (and similarly, sufficiently powerful personal computers with resources to spare), if a site presented a series of images and the user wanted to keep them, they would either have to laboriously click and download each image or download the entire page to hard disk and copy the image files, or write a bot to do so. Now additional browser software can provide automatic downloads, even allowing users to specify the media type (pictures, videos, sound files, documents etc) and - crucially for the context of this chapter – explore all of the links on the target website in order to find all of the requested media across the site. Such a task would have taken a painstaking number of hours to do manually if not automated. This would have been especially labour intensive in the early days of the World Wide Web if dealing with a website that had more than a handful of pages and desired media items (e.g. pictures) numbering in double figures or higher on each page

[45] This is no longer an insignificant concern. Hosting provider Incapsula analysed a substantial portion of its hosted sites' traffic and found that 61.5% was originated by bots. (Incapsula, 2014)

Programming Interface) is a structured interface or method for obtaining information from the site programmatically and is generally preferable to having to write bots that are able to read the site's unique structure manually. A site's RSS or Atom feed (an automated list of new site content or output) can be regarded as a very simple API, for example. That being said, an API can still be limited – not just in terms of the number of queries it may permit at any one time but also the type of queries. Information could still be hidden that is not accessible. A common task for bot-makers is to write a spider that iterates through the user accessible forms on a site as they often contain drop down menus with additional information and may be the only way to access one of the site's databases and glean information that isn't necessarily available either through a straight forward screen scrape or an API. This specifically requires imitating a human. This raises ethical issues regarding identity and potential fraud online[46]. The Python module 'Mechanize'[47] for example is designed for automated form filling in on a web site and subsequent capture of output – it is specifically tailored to allowing code to identify itself as a particular browser rather than a bot. In contrast to the robots.txt file, details of an API are usually found on a 'terms of service' page, so the onus is on a human rather than a bot to identify what facilities are available already on the site for extracting information.

An API is also desirable for both bot writers and webmasters because it, in theory, should shield any bots that use it from underlying changes in the way the site is structured or output. The HFT bots that were the major catalyst for the 2010 Flash Crash would interact with the APIs for each exchange or market, for example. The fact that knowledge regarding the API protocols was not universally shared or understood was definitely a contributory factor to

---

[46] There are numerous mechanisms that are used ubiquitously such as CAPTCHA (see http://www.captcha.net/) in order to tackle the identity issue, putting a roadblock in front of a bot for a service that is intended for human consumption. However, they have to constantly evolve as bots become more sophisticated.
[47] See for example here – 'Emulating a Browser in Python with mechanize' - http://stockrt.github.io/p/emulating-a-browser-in-python-with-mechanize/ or here – 'mechanize' - http://wwwsearch.sourceforge.net/mechanize/

both the growth of particular HFT businesses and to the chaos that ensued during the Flash

Crash (Patterson 2012; Steiner 2013). The chains of Circulating Reference were easily

broken because some of the intermediaries were Black Boxes that were only opaque to

certain parties. APIs also customarily remove the need (speed aside) for some bot writers to

shield the fact that it is a bot interacting with the site. One of the surefire ways for a

mechanizer bot to be identified for example is if it submits form data (by imitating the form

intended for human users when sending its request to the server) that would have been

impossible for a human user to generate with the current form[48]. Most HTML and/or

javascript created forms online have validation features built in that may be invisible to a bot.

Interfacing via an API creates a structured environment that prevents this kind of mismatch.

Similarly, *screen scrapers* are commonly 'broken' by behind the scenes changes that would

not be noticed by a human browsing the site but matter a great deal to the correct functioning

of a bot, especially if it is already locating the relevant items within an assumed structure or

in a particular order[49]. Paul Bradshaw, for example, in his guide for journalists who want to

make useful bots 'Scraping for Journalists'[50] discusses how to deal with a particularly badly

formatted page on the Labour Party website[51] that contains a list of meetings journalists may

be interested in. He contends that formatting such information in a feed or making it

accessible via an API might have been much better for all parties concerned. That is, by

---

[48] For example, by submitting an incorrectly formatted date if the human facing interface has some kind of date data validation in place.

[49] The offending change can be as minor as an alteration to the name of the element being searched for in the raw text (HTML) version of the page. This can easily occur as a result of a third party without the webmaster even being aware, highlighting once again why close empirical analysis of such worknets are important. For example, a site may rely upon advertising, which is placed within a designated section of the website. This content is fed through directly from a third party. If the spider is designed to reliably skip the entire section it is possible a different kind of advertisement, with a different format, could appear and break the spider if its algorithms rely on identifying a specific format to skip. Such a change would this time not only be invisible to the end user but also the webmaster. It may cause severe problems for a spider however, especially if the page output is not clearly formally defined and structured

[50] (Bradshaw, 2013)

[51] 'Ed Miliband: List of meetings and dinners with donors and trade union general secretaries' http://www.labour.org.uk/list-of-meetings-with-donors-and-general-secretaries,2012-03-30 – last accessed 24/09/13

making the Circulating Reference chain more 'rigid'. APIs typically lack the same kind of restrictions laid down in the robots.txt file – or at least any restrictions tend to be more generous recognising that it is generally automated processes that will be making use of it and thus leveraging the speed of a bot rather than the plodding pace of a human enquirer.

A general account of Bot description

There are many behaviours and activities that have comprehensive customary codes, legal regulation and normative expectations attached when accessing networks through browser intermediaries that are explored, for example, in (Fuller 2005, 2006; Solove 2008; Benkler 2006; Bucher 2012; Rogers 2006). These can all be called into question or even made redundant when custom bots are created to carry out tasks that are typically carried out with a browser. Although standard practices and conventions have been developed and often adopted in the use of bots in this way, actual practice in the field guarantees a plethora of ethical and legal grey areas. The grey areas in question are useful for highlighting the complex relationships these bot technologies are deployed within, not least via demonstrating that a correct understanding of their use in practice requires some kind of active empirical description. If using Latourian notions for this task, humans can appear at any point as actors or actants – i.e. those *acting on others* and those *being acted on* respectively (Latour, 1993b, 2005).

The use of such bots, in searching out information online, reflects a return to the early days of both the internet and networking more generally where scripts and command line instructions were the norm for any communication, exploration or action. They bring one effectively 'closer to the machine'. The ubiquity of the browser has served to Black Box many of the underlying processes that still occur and in most cases, operate on the basis of the very same

protocols (e.g. FTP, HTTP etc.) that are a browser's bread and butter. The fact that the user's experience of the internet presently is usually a browser-mediated one, rather than a command line, is one of the main contributory factors to fuzzing the edges of ethical and legal bot deployment, creating significant uncertainties and ethical vagaries – not least because legislators and policy makers themselves primarily consume the internet through browser intermediaries and thus understand it on that basis.

Bots can behave in unconventional ways that typical browsers usually won't, or at least browsers that have not been compromised by some kind of malware. As a result, bot activity on internet sites and networks may automatically be viewed with suspicion and easily misinterpreted if it is noticed. Given the surfeit of add-on functionality for modern browsers however, it may not always be so easy to make the distinction. With this in mind, for current purposes I make a distinction between 'off the shelf' browsers and bots. Browsers enhanced with particular add-ons bear capacities more similar to bots and as such are subject to the same considerations (and classification) as the latter. A typical off the shelf browser works primarily as an intermediary for a human to access the internet and little else, with its primary activity to simply translate the underlying website markup into something that is ordered into something comfortably comprehensible to humans. This is of course what use interfaces (UIs) of all descriptions, not just browsers, are fundamentally designed to do[52].

These kind of browser substitute bots (encompassing *spiders*, *scrapers* and *mechanizers*) illustrate the utility of describing them as parts of assemblages in Latourian terms, and in particular in *worknet* terms[53] that include humans because they "perform based on their changing environments, and they respond slightly differently every time they run. As a result,

---

[52] See for example, (Galitz, 2002).
[53] As described in Chapter 2, a 'work net' is an alternative name for an actor-network proposed by Latour to distinguish it from current computer networks. The two are not necessarily synonymous. Also the shift in terminology emphasises the 'work' element in that for anything to meaningfully (and observably) be a worknet (actor-network), its various actors and actants have to do *work* that leaves behind *traces*.

webbots become capricious and lifelike" (Schrenk 2012, p. 11). Moreover, effective use of spiders, scrapers and mechanizers is predicated on both working around browser limitations as well as capitalising on them with regard to human users. For example, a business model based on price comparisons has a distinct advantage if based upon bots that sought out and processed the relevant information if competitors used just human researchers, further slowed by browser intermediaries, to achieve the same task. By way of contrast, the 'newsnow' website is limited by its bots' adherence to the Sun and Times' directive files discussed above, leaving human users to seek out the latter's content manually. It is also important to recall that the human users may not even be aware that they are missing out on content from these sources[54].

What is or is not 'appropriate' behaviour may change according to whether the actor/actant in question is a human using a browser or a bot acting autonomously[55]. Thus the worknet any one piece of code is situated within matters as much as the capabilities and potentialities (i.e. decision making trees, situational access) of the source code itself. There is a traditional nomenclature for bot programming and deployment that refers to their being explicitly *launched*. That means they are either executed manually, or scheduled via a service such as ScraperWiki[56] to work at specific times and have one or more external targets (as in, usually external to the network from which they are executed). If the bot is being explicitly activated, it is effectively foregrounded and thus more amenable to accurate description and analysis. A bot that acts as a function behind a web interface, callable by any user, remains in the

---

[54] I certainly didn't realise myself despite having been a long time user of newsnow until I inspected the Sun and Times robots.txt files.

[55] Whilst it is not within the scope of this thesis to go into any great depth on the many normative questions raised in studying bots and networks, forays into their ethical lacunae can be sharply illustrative of the concepts and issues I am attempting to articulate here. Ethicists' attempts to define and situate concepts that encompass the novelty that is characteristic of digital object(s) of study and the problems they raise provide significant insight. They bring us back time and again to the underlying ongoing problems of describing dynamic socio-technical systems and premature (possibly unnecessary) Black Boxing

[56] https://scraperwiki.com/ - a cloud based service for hosting and launching multiple bots.

background, effectively Black Boxed[57]. All of these different methods and (ab)uses of bot technologies can come together in all sorts of configurations. Thus attention to the context becomes critical.

Identity, automation and expectations

The 'User-Agent' header usually provided by any browser agent to the target web host is specified as a field in the robots.txt file for blocking blacklisted or affecting otherwise limited bots. Similarly, the webmaster relies upon it in the logs for correctly identifying where the site's traffic is originating from. This header is stipulated in a number of the robots.txt files detailed in Appendix 2 in order to name particular bots and instruct them to behave in a particular way (usually to not index the site at all). It is however an optional header in the HTTP protocol, making it problematic to appeal to as a fundamental feature in determining ethical issues, never mind policy responses attached to 'correct' conduct online. The only guarantee that the entity accessing the site is who they claim to be in the 'User-Agent' field is practical considerations regarding what they want from the site. It may be necessary to inform the site that a particular browser is being used so that it sends the correctly formatted page – a crucial consideration if browsing from a mobile device for example. As (Rhodes and Goerzen, 2010, p. 152) put it: "you can simply instruct it to lie about its identity, and the receiving web site will not know the difference."

---

[57] Bots that behave more like functions may be executed by users with little to no awareness of what is occurring behind the scenes, such as is the case with a user's experience of using Google – an apparently relatively straightforward function that is in fact dependent on an enormous amount of bot activity in the background - see for example, (Rogers, 2000, 2006). On the other hand, a bot that exists in source code on a hard drive could easily be called by another program at any time and executed in ways that may not have been foreseen by the original developers. A simple *mechanizer* that was intended to query the contents of a single page occasionally can easily be turned to other uses, such as being called repeatedly as part of a denial of service (DOS) attack. And to return to the Google example, there is a subset of hacking practices known as 'Google Hacking' that leverages the search engine's indexing services for unintended purposes (such as exposing personal information, accounts and so on).

Even though many users may still be unfamiliar with the notion of an internet beyond the browser, the exploitation of this gap by bots is now becoming less of a novelty for research given the steady encroachment of bot and Big Data technologies into further domains that were until only recently, the preserve of humans. Many examples are detailed in (Steiner, 2013) and (Lanier, 2011,2013). Their examples highlight how it is now a matter of course to capitalise on browser limitations. Steiner highlights cases such as classifying algorithms that can grab information freely (though unwittingly from the user's point of view) shared with servers by the browser. Similarly, from the other end of the interaction, Lanier explores how browser data can be used by social media services to sell both users' data to third parties and place targeted ads. A *spider*, *scraper* or *mechanizer* can identify itself arbitrarily as a browser, any browser, rather than a bot. Similarly, the 'Referer' setting/field provided by the bot (or browser) can be set to any arbitrary value and used (amongst other things) to download content that would normally be prohibited from being shared beyond the site, usually in order to preserve bandwidth. Images are often 'hijacked' by other sites for their own purposes – e.g. dating websites may steal images to create fake profiles in order to give the appearance of a more active site. These are capabilities that a typical browser equipped human would not be able to exercise – and part of the power exercised by the bot (on behalf of their user or indeed, other bots that may launch them) is actively deceiving the target website into treating it as a browser and similarly, using the 'Referer' trick, pretending that the link to the requested data is coming from elsewhere.

What kind of consequences can these grey areas have in practice? In visiting a website or similar facility, whether as a human with a browser or a human instructed bot, how one expects a particular object or technology to behave depends upon the expectations already in place before it is even deployed. Crucially, such expectations can be heavily context dependent. Johnson's work, for example, illustrates this particularly well. She and her co-

authors consider the issues arising from the digitisation of campaign finance disclosure in the U.S. (Johnson, Regan and Wayland, 2010). This data was originally made publicly available in the interests of transparency with regard to who the donors may be for any particular candidate or party. However, as Johnson et al demonstrate, its subsequent digitisation led to numerous unintended consequences such as removing "practical obscurity"[58] for those who were subjects of the data. These are very similar to concerns raised over the potential of 'Google Hacking'[59].

Originally some effort would have to be expended in order to analyse the donor data. However, once made available in a complete and digital format, second order data users were quickly able to make new uses of the data; many of which had privacy implications – especially for donors themselves. The motivations of those seeking the records rarely mattered when following the more labour intensive procedure of checking the records by hand – thereby guaranteeing the "practical obscurity" provided by such effort. However, when the entire data set is made open to the digital wild with a negligible bar for effort, the motivations involved can matter a great deal, thus invoking much more urgent ethical questions[60]

> The interested viewers may have benign, nefarious or legitimate reasons for being interested, and their interests may or may not be related to the original purpose for making the information public. (Johnson et al, 2010, p. 959)

---

[58] This refers to the fact that, prior to digitisation, there is a requisite amount of effort required to gather this data. The effort required means that curious people are less likely to investigate, thus giving the subjects of the data 'practical obscurity'. Once it is digitised and online however, the information is available quite literally at the flick of a switch.

[59] 'Google Hacking' is the practice of using Google's numerous search facilities in a way that imitates hacking in some respects. It can be used, for example, to find information that has been unintentionally shared, active webcams and so on.

[60] These types of issues are addressed directly by many scholars in the area such as as (Solove, 2008), (Morozov, 2012,2013), (Lanier, 2011,2013) and so on.

The possible break with the "original purpose" here is key. In Latourian terms, the data set has been placed within an entirely different worknet to that originally conceived. And along with other actors, actants and worknets that necessarily act as intermediaries and transform the data in new ways. That a technology (or data set) in singular or general terms can be transplanted like this, and often so easily, begs the question for technological determinists as to a technology or particular tool's inherent capabilities and destiny to act in the world. The change of relationships, and therefore context, changes with it the values that may or may not apply. One such transformation in relationships Johnson et al identify is the method of copying – using the 'legacy' method, a researcher would have to copy the entry by hand at a public building. Now every single record can be copied effortlessly and shared or sent elsewhere. Moreover, selected parts of data can be removed and reconstituted with other data sets, yielding not just further privacy concerns, but also issues of misrepresentation. The donor data is already largely separated from important contexts (such as giving no clue as to the donor's motivations) and could be further distorted by being placed in new contexts that may have little or nothing to do with the original decision to donate. This is part and parcel of the current design and infrastructure of networking technology. Though as (Lanier, 2013), (Zittrain, 2009), (Morozov, 2013) and (Galloway, 2004) all point out, *it didn't have to be* as alternate protocol designs are possible. The original network engineers may not have been able to envision the conundrums that almost effortless copying would lead to, however it was specifically a capacity embedded in the design and most certainly not driven by the technology itself. Alternative models of networking *could* have been developed. No digital configuration is set in stone, it may only seem to be so because of the practices within which it is embedded.

Johnson et al. summarise this constellation of concerns surrounding (easy) copying as a matter of intensifying the tension between privacy and transparency. There are numerous

political angles to this that the authors explore which I will not rehearse here[61]. What is relevant however is how bots are able to facilitate these unintended consequences and may in fact frustrate efforts to rebalance the privacy and transparency compromise that are based upon the assumption that whilst the new digital format of the data makes it highly portable and recombinable, the main culprits for use or misuse will be humans armed with nothing more sophisticated than a spreadsheet or database application with which to sort, categorise and modify the data. Bots remove almost every 'human level' restriction on speed or labour where movement, analysis or transmission of data is concerned. They also enable rapid changes in context and meaning as (Kirilenko et al, 2011), demonstrated – the bots' behaviour could change very quickly so that they may be reclassified as belonging to different market participant groups engaging in significantly varying behaviour at numerous points.

Are bots saving human effort, or increasing it?

Steiner (2012) presents a compelling narrative detailing how algorithms (bots) have entered many aspects of life and displaced the humans who were there – from trade floors to the creation of music. The logic has inevitability about it – as soon as a particular task or skill is representable in reliable algorithmic form, it is only a matter of time before it is able to supplant human labour. Moreover, in some domains, the bots can actually outperform the humans they replace. Examples abound in Steiner's account such as chess playing or imitating the musical style of a past musical master.

Even in domains where humans cannot be replaced for now, bots can be deployed that act as a kind of cognitive or physical force outsourcers – as per Wiener's anticipation of widespread

---

[61] Which include, for example, an artificially greater scrutiny of candidates that was not practical before without significant effort. As Johnson et al explain this has had effects such as forcing candidates to suddenly have to explain donation patterns that were previously given scant attention and were effectively normalised.

delegation and increasing concern of 'machine takeover' by default. This could include mass delegation for ethical decision makers (Baase 2007; Singer 2011). This is already occurring in, for example, medical ethics in practice and drone warfare in the Middle East respectively and a crucial endpoint where almost everything is automated is envisioned from this trend (and railed against) by (Lanier, 2011, 2013).

The findings from the robots.txt files demonstrate that human (and to some extent non-human) actors outside the virtual space form a key part of the total networked assemblage when considering bot behaviour. This is especially apparent relative to the directives included in a robots.txt file. We should go beyond discussion of software just in terms of a form of literature and must similarly attempt to go beyond attempts to impute the intentions and thinking of the original programmers from careful study of the style of coding, commenting and so on. It could be risky to draw firm conclusions without the wider context when carrying out a text based analysis of source code. This is why one risks instant censure in most coding workplaces for not including appropriate commenting for the benefit of anyone else who later has to maintain one's code – such as in the amusing Microsoft example discussed by (Berry ,2011, p. 69), where code maintainers are threatened with 'being killed' if they modify a specific piece of the code.

One of the reasons it is so important to catch the dynamism at work *in situ* is that it is a fundamental aspect of the technologies under the microscope in a way thatwas not the case with previous technologies, no matter how 'revolutionary'. Floridi, for example, makes a case for arguing that computing technology and the ethical theories (the 'macroethics' theories) mark a fundamental break with the past history of ethics in philosophy:

> ICT [Information and Communications Technology], by transforming in a profound
>
> way the context in which some old ethical issues arise, not only adds interesting new

dimensions to old problems, but may lead us to rethink, methodologically, the very grounds on which our ethical positions are based. (Floridi, 2006, p. 23).

The typical view advanced on macroethics[62], Floridi argues, dismisses the philosophical significance of both ICT and a specific branch of philosophy dedicated to it, holding that it is at most a microethics within larger, already well-established macroethical structures. Following the Information Ethics(IE) directives, Floridi argues that they deal not so much with 'right' or 'wrong' as what is actually better or worse for the 'infosphere' (the totality of all things understood as reducible to a fundamental medium – information). In a macroethical sense, for Floridi, contributions such as IE and ANT provide distinct extensions and additions to the pre-existing ethical frameworks (not to mention epistemological, ontological and so on).

Whilst Floridi does attempt to consistently stretch this conception to all entities – some (myself included) do not find this comprehensively or consistently defensible on the grounds that Floridi's information directives are not necessarily universalisable, for example, (Stahl ,2008). This is less problematic though if the domain of study remains purely within the virtual and digital ontologies of ICT – i.e. on a flat digital plain where human interactions are represented only digitally by the computing substrate. On those grounds, as an ethical tool, IE is tremendously useful and consistent when referring to data entities and the agents that interact with them *if talking in purely virtual or digital terms*. As Sara Baase argues in 'A Gift of Fire', with ethics in general and the ethics of technology in particular, it is best to pick the right tool for the job in hand (Baase 2007), rather than look for a single universalisable macro ethics theory. As a shorthand for deciding what to do, in ethical terms, IE is excellent

---

[62] This is the view that computing and information technologies are just adjuncts to existing technologies. As such the ethical problems to which they give rise are similarly viewed as falling within pre-existing domains (e.g. Deontology, Consequentialism, Aretaic theories etc.) rather than demanding a new category or approach in their own right.

in the purely digital realm. Unfortunately tracing worknets related to any particular

controversy is unlikely to remain so limited, therefore restricting IE's useful applicability.

IE's list of directives focus on maintaining the integrity[63], and if possible, improving the

quality of information in the 'infosphere' with which one is dealing. The primary locus – and

test - of ethical standards here is to consciously avoid imposing entropy (loss or corruption of

information) and indeed to minimise it where possible. Slotted within a Latourian view of

technological and informational objects then, this provides a candidate for normative

direction that is lacking in Latour on his own, even if (just like Latour's approach) it has to be

worked through for each unique situation and worknet and (or 'infosphere') in question. A

Latourian analysis provides an excellent framework for describing and conceptualising what

the object(s) of study is. IE could then be applied to understand what agents should then be

able to do within it, at least with regard to the digital aspects. Latour is quite clear that he is

relaxed about the 'normative shaped gap' (Latour, 2005) and does not consider its absence as

a fatal limitation for his approach(es). It is however a component that needs to be inserted at

some point for wider discussions beyond merely the descriptive or epistemological and IE

provides a compatible 'plug in' for this kind of purpose, though other accounts are of course

entirely possible. Both however, provide additional filters or contrasts with which to

foreground the bots and Black Boxes that we want to make visible and open, respectively.

In contrast to IE's clear normative thrust, Latour frequently reiterates the amorality of ANT

as an *observational method* (Latour, 2005). Generally, some kind of civility is maintained

between 'bot' writers and webmasters (or network administrators) by following informal

rules of 'netiquette', of which the robots.txt file is a key part. However, the inferred intent of

---

[63] Those directives being, from (Floridi, 1999): - 1. uniformity of becoming, 2. reflexivity of information processes, 3. inevitability of information processes, 4. uniformity of being, 5. uniformity of agency, 6. uniformity of non-being and 7. uniformity of environment.

program writers by a human or machine agent on the other side can make or break this relationship between the two groups.

With the spate of Denial of Service attacks in 2010 on numerous websites by the hacker group Anonymous[64], the intent was very clear – and the attacks (mis)used a characteristic of the underlying networking technologies to achieve their goal[65]. And yet perfectly legitimate data-gathering requests could be sent by another party using a clearly identified 'bot' that sends requests only infrequently in a situation where the server happens to be under a large amount of strain. If the server subsequently crashed or stopped responding, some webmasters may regard the 'bot' activity as hostile. A fairly modest rate of querying (say once every three seconds) may be seen as onerous relative to the remainder of requests that the server is receiving on a particularly busy day.

Such considerations apply to the 'bots' for the 'churnalism' research described in Chapter 5. They were specifically launched at times when the servers in question would be under less strain and maintained a choke of one request every three seconds. Despite this, there was a response from the webmasters of both the Environment Agency and DEFRA websites. An attempt was made to block access from the IP address used to the sections of the website that contained the press releases. Under common 'netiquette' for 'bots', my bots were being classified as 'bad' 'bots'.

The entire situation must be regarded as an 'assemblage' of the webmaster, the server, the 'bots' and the programmer launching the 'bots'. As highlighted earlier in the chapter, the 'meaning' – and even one might say, the 'identity' - of the 'bot' now only becomes clear in this totalising context. The 'bot' cannot necessarily be identified beforehand as a badly behaved program without the context of the server it queries and the state that is in, along

---

[64] See for example, (BBC, 2010) , *Anonymous activists say Wikileaks war to continue*.
[65] By flooding the target servers with SYN/ACK requests, something that can be done at very high speeds.

with the competing priorities of the webmaster which appeared to be in this case, a realisation that the 'bots' were scanning every single press release and an, as yet undisclosed, reason for stopping it[66].

By way of comparison, when the press releases were submitted en masse to the Churnalism site (churnalism.com), there were no attempts to restrict access (see next chapter). It is also worth bearing in mind that the site received approximately twice as much traffic from the bots as data for the target sites were both submitted to this one and in a shorter space of time. This presented an entirely different assemblage and indeed the team behind the Churnalism.com facility have expressed their desire previously for people to make as much use of the site as possible[67].

Summary – ground work for better descriptions

Using the robots.txt files as source material I have laid out the various narratives and factors that come into play in the interaction of browsers (human or bot), websites, servers and webmasters. These numerous observations illustrate the delicate threads that can make up a chain, or chains, of Circulating Reference and how easily they are interrupted or broken. One particular salient factor in their description however is how, unlike the trail of Circulating Reference Latour describes for the soil researchers, it can be easily repaired and set in motion again, often relying on the dynamic interplay between the different components involved.

---

[66] It is worth consulting the (Brooke, 2011) work on this, particularly chapter 5, on how similar efforts have been stymied in the past such as using publicly available data to create 'the public whip'website (http://www.publicwhip.org.uk/). Plus of course the concerns already discussed in (Johnson, Regan and Wayland, 2010).

[67] In one email exchange I had with the webmaster, asking how quickly and at what times he would prefer my bots queried the site, he responded, "In terms of hitting the churnalism.com API, I say just go for it. If you notice it slowing down, crashing or letting out smoke, let me know ;-)".

This interplay discussed above between a site visitor (bot or human), the site/server itself and webmasters provides a useful accounting of a digital-centric sociotechnical system from which some insights can be gleaned.

Firstly, it is clear that intentions, even if clearly coded into a legible form whether as a directive file (robots.txt) or within the code of a bot are usually an insufficient basis for understanding what might go on when such entities do end up interacting 'in the wild'. This point may in fact be best characterised in the same way as the robots.txt file itself – it gives an indication of how its creators or users would *prefer* events to play out and yet in practice is something often distorted, or only partially followed if not outright ignored. Take the British Chambers of Commerce file for example. Bots would almost certainly ignore the 90 second delay directive even though, had they been present, they may have honoured directives to not index particular directories and files.

Secondly, deciding 'what must be done', whether the normative focus is ethical, political or just practical, whilst an interesting, even critical, area of concern is automatically pre-figured by any and all efforts to describe the system in question in both ontological and epistemological terms. Whilst there are obviously always going to be enormous unknowns inherent in such questions in the still largely heterogeneous internet[68], where there is any kind of regular event, system or process in play there are relationships that can be described using or adapting several Latourian concepts. The *relationships are key*, relative to the computational possibilities and degrees of freedom of the source code itself. The primary step then is in sketching out relationships, and where possible intentions, something that analysing robots.txt files can be useful for in client-server web contexts[69]. Circulating

---

[68] e.g. the potentially infinite range of alternate circumstances, variable inputs etc. that a digital or human actor/actant may encounter.

[69] Though redundant in practical terms for the 2010 Flash Crash beyond considering the wider digital ecosystem of information feeding in from the 'outside'.

Reference is clearly well suited to this task – tracing the provenance of references made in the system to other entities and states of affairs. That is not, however, the most critical reason for introducing this concept and updating its application for the digital era.

That reason is found in the third major point of consideration raised by analysis of interaction via the robots.txt files. That is the sheer delicacy of some of the relationships at work. It is no good describing these relationships as they are found if they are not also characterised with a relative level of strength or reliability. This is something that Circulating Reference is able to do along with related Latourian concepts such as Trials of Strength and Obligatory Passage Points. The chain of reference used for the initial research outlined here was itself broken part way through in two points. One being critically reliant on the algorithms at work in one particular Black Box, (the 'churnalism.com' facility), which suddenly changed to an incompatible state. The other was changes within one group of the candidates for observation – the government departments. It doesn't really matter what we believe *should* or *should not* obtain. What we believe is known, or can be known as well as how we believe something is structured necessarily underlies any decision making, or even decision making capabilities. This is an issue in terms of establishing just what *can* be done; never mind what *should* be done. In a networked world, the relative strengths and weaknesses of ever changing connections and relationships are important for understanding this. We want to understand not only what is grasped, but also how strong the grip is and whether it can cope with every dynamic, or *elastic*, relationships and referents. Circulating Reference, as useful as it is, remains backwards looking and post-hoc. It requires additional components; one may even say help, to operationalise it as a tool for describing present and future scenarios. I now turn to consideration of more dynamic subject matter (automated churnalism detection) and Computational Coherentism in order to make clear what kind of gap has to be filled for Circulating Reference to work in this context and to assist in the overall task of how to

apprehend fast moving networked assemblages.

Chapter 5: Case Study: Churnalism

This chapter proposes one possible group of approaches for apprehending fast moving networked assemblages that act in a genuinely heuristic fashion and complement the Circulating Reference heuristic already recommended. They also enable a closer to real-time analysis and are highly adaptable to changes by design. I begin by sketching a real-world problem: 'Churnalism'. This phenomenon has had a significant effect on our ability individually, collectively and institutionally to assess both the provenance of information and how to choose between apparently equally provenant data (Davies 2009; Lewis, Williams and Franklin 2006). The practice of churnalism impacts news data that has historically had significant effects on market movements, it is also therefore a concern in and of itself regarding events such as the Ohio power outage and 2010 Flash Crash. Conflicting narratives with and without varying levels of provenance are a significant problem when attempting to describe what is happening and what has happened. Dealing with this kind of difficulty requires an approach that maintains the ability to adapt descriptive heuristics to dynamism in the same way and move similarly flexibly with the ever changeable epistemology and ontology of one's object(s) of study. I propose and critique *Computational Coherentism*, specifically Paul Thagard's flavour of it, as a strong candidate for progressing the inquiry in this direction.

The research I carried out used the pre-existing 'churnalism' facility[70] in an automated fashion, leveraging its API to fire off hundreds of queries in quick succession. It yielded data in which there were discernable patterns. Moreover, the data also challenged some of the assumptions in the original (Lewis, Williams and Franklin, 2006) research carried out at the University of Cardiff that underpinned (Davies, 2009). The Cardiff research was the first

---

[70] Churnalism.com

empirical analysis of how PR releases and news wire articles were shaping the output of the

UK media. This was contrasted against a background of journalists being expected to do

more with less time[71]. On the basis of the two week sample of output analysed by their study

it was suggested "that 60% of press articles and 34% or broadcast stories come wholly or

mainly from one of these 'pre-packaged' sources" (Lewis, Williams and Franklin, 2006, p.

3). They go on to note that even though the broadcast media apparently used less 'churned'

PR content, it nevertheless had a detectable influence in "playing more of an agenda-setting

role."

An oft repeated caveat in their analysis is that it is inevitably conservative in its estimates for

the influence of PR (PR here standing in for 'Press Releases', though 'Public Relations' is

also apropos) on the media. Something confirmed by my own findings. PR copy is often

cleverly paraphrased and is also, as the Cardiff study notes, often the basis of many wire

agency releases that are themselves treated subsequently by journalists as objective work.

Furthermore, attribution of sources by the media outlets themselves is often obscured – for

example:

> "in nearly a quarter of cases there was no clear indication who had written the
>
> story…Only 1% of stories were directly attributable to PA [Press Association] or
>
> other wire services…At first glance, then, these data suggest that the newspapers give
>
> the impression that they depend on their own journalists rather than wires or other
>
> outside sources" (Lewis, Williams and Franklin, 2006, p. 15).

The Cardiff analysis covered the output of the Guardian, Independent, Times, Telegraph and

Daily Mail. For broadcast media it analysed BBC and ITV evening news bulletins, Radio 4's

---

[71] A neglected area here is also the details regarding how journalists and journalism are adapting to new technologies and without 'technological utopianism' that regards it as some kind of fait accompli or manifest destiny – something which has only recently started to be addressed e.g. in (Schmitz Weiss and Domingo 2010; Bowker et al. 2010).

'World at One' and the 'Today' programme. It covered a total of 2,207 press stories and 402 broadcast items. Each item was painstakingly researched to identify its original sources – an extremely time consuming (not to mention costly) activity that provided one of the inspirations for my own research into developing ways to similarly assess media content in an automated fashion.

They opted to focus primarily on the 'quality' press, on the basis of the assumption that they were "the places where we would least expect to find a reliance on pre-packaged news" (Lewis, Williams and Franklin, 2006, p. 13). My study disputes this assumption, finding instead that the 'quality' press were actually the worst offenders for cutting and pasting content from press releases from the two organisations selected for analysis, with only one tabloid, the Daily Mirror, even noticeably producing substantially 'churned' content.

My own approach essentially reversed the methodology of the Cardiff research. This was necessitated by the (at the time) way the 'churnalism' facility operated itself. Instead of starting with the media organisations it began with the press releases of two organisations – DEFRA and The Environment Agency. The main rationale for this approach was that the Cardiff research indicated so many difficulties in tracing 'churn' back to its original source and potentially introduced a lot of subjectivity in analysing ambiguous content. Furthermore, they repeatedly point out that the identification of government PR in the media is particularly difficult – at least it was using their methodology of starting from the final product. I therefore selected two government agencies for the initial study, with a view to adding many more at a later date (the originally planned cohort is described in the previous chapter). In any case, any subsequent round of research using this methodology should cast a wider net over a larger range of PR sources. At the time the analysis was carried out, DEFRA and the Environment Agency had both recently produced articles that had been widely (and

uncritically) repeated by the print media[72] and they also concerned environmental issues that were often deemed of public interest by the media and so routinely reported upon the basis of government press releases. It is probably helpful at this point to be able to visualise how 'churn' can be depicted. An example of manually inputting one of these stories into the original churnalism facility is given below:
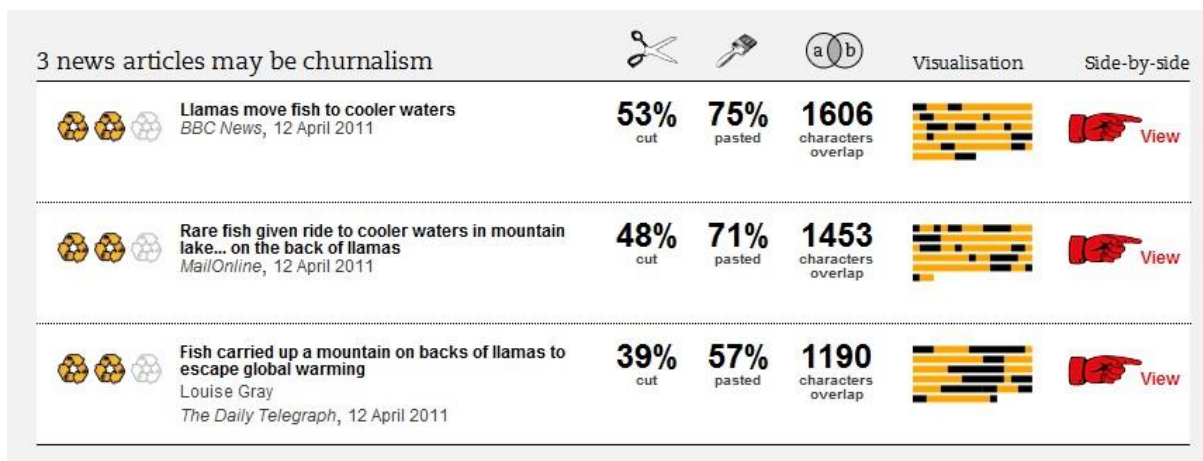
Starting with the original PR output also meant that more interesting questions could be asked. The Cardiff team provided useful bulk information on output that was very simply categorised – i.e. whether the originating source appeared to be businesses or NGOs. I hypothesised that in any follow up research that specific patterns of editorial bias could be detected, essentially revealing clear relationships between particular outlets for press releases and the media organs that replicated their content and, further, at a more granular level of detail, individual journalists' use of PR material in these instances could in fact be identified.

---

[72] From the Environment Agency 'Llamas help protect ice age fish' and from DEFRA 'New service for householders to stop unwanted advertising mail' – see the online repository: https://www.dropbox.com/sh/qbufxoi0j85vik7/AACW7nvVWMqjBT0qPmGKgzY9a?dl=0

This hypothesis is borne out somewhat in the modest amount of automated churn detection carried out, though the overall approach should be applied to much larger datasets to confirm this.

The range of outlets for PR material is huge, so any kind of representative sampling for such a modest study was out of the question. The aim was, instead, to demonstrate the practicality of automating the study of PR content with a view to crowdsourcing a much larger effort in the future that could provide a large amount of coverage across the whole range of issues and political spectrum represented in the UK media. Computational Coherence algorithms could then be used to sort through the narratives proffered. My research also showed that, in theory, a near real time coverage is possible that would enable consumers to make much more informed decisions about how to interpret mass media content in the press as well as for traders and bots engaged in making investment or divestment decisions.

Press releases were gathered automatically by web spiders written in Python and then submitted to the Churnalism engine (see Appendix 3 for the main script code). From the DEFRA website, a total of 386 press releases were analysed, from 13th May 2010 to 24th November 2011. For the Environment Agency a total of 1962 press releases were analysed from 27 January 2010 through to 17 November 2011. The difference in dates (and numbers) reflects the total availability of press releases on either site. Each press release was then submitted to the 'churnalism engine'. The facility compares any input to the content of all articles contained in the 'journalisted' database (a database that indexes most of the online output of UK national newspaper websites) and provides a direct 'cut and paste' comparison and 'churnalism score'. The bots created to submit scraped content from the PR outlets to the churnalism site interfaced with the provided API. Returns from the churnalism.com API give a precise 'score', calculated from the proportions that have been cut and pasted respectively.

A three tier scoring scheme was then applied to the results: 1) "detectable" churn - with scores of 100 or more (in practice this would mean maybe a paragraph had been copied) 2) "significant" churn - with scores of 500 or more (in practice more than one paragraph had been copied) and finally 3) "major" churn - a score of 1000 or more meaning the majority, or substantive minority of either/both press release and final article had been copied. The most striking patterns were as follows: For the Environment Agency, setting the filter on the data I gathered to a score of >=100 ("detectable" churn) yielded 5089 articles in total. A clear pattern emerged with the BBC, Guardian and Independent producing the most 'churned' content in the higher churn categories. 1983 of these were from the BBC, and out of those 174 fell into the "significant" category. The relative percentages of churn across all three categories were sufficiently close to demonstrate a definite pattern. A very similar result was obtained with the DEFRA data set, which generated 1959 cases of "detectable" churn. One of the most surprising results was, as mentioned earlier, the relative invisibility of the tabloids – especially in the higher churn categories, with the Mirror being the only one to reach double figures (10 counts) in "significant" churn of DEFRA press releases.

The above represented a "simple" methodology that could be built on in later iterations. As the Cardiff study details, many articles are in fact made up of several press and wire releases, meaning the actual amount of churn is much higher than even represented in the above data I gathered as it was in the manually gathered data in the Cardiff study. Subsequent research should adopt a 'multi-pass' methodology where results from each individual data set are collated to find duplicate articles and the results then combined to provide a more accurate analysis where articles have been 'churned' from more than one press release. This does not help epistemologically. That is the task for Computational Coherentism in the second half of the chapter – it does however at least separate the multiple sources of narratives in these cases.

The dataset returned from just 2348 press releases was substantial – returning thousands of data points from the churnalism engine. There are likely other interesting patterns and links yet to be found in the data. The data has also been made publicly available online should anyone wish to replicate the findings (or indeed, find other interesting patterns)[73].

What kind of effect can such 'churned' information have in the wider world? Davies (2009), referring to the events leading up to the New Years' Eve of 1999 and panic over the 'Millenium Bug' highlights an important example where it becomes crucial to determine the provenance of information and how far out its influence radiates:

> Journalists reported that the British government had spent £396 million on Y2K protection. They also reported that it had spent £430 million. And that it had spent £788 million. The American government had spent far more, they said - $100 billion, or $200 billion, or $320 billion, or $600 billion, or $858 billion, depending on which journalist you were reading (Davies, 2009, p. 12).

Adopting a charitable interpretation, as Davies does, we can probably assume that most of those figures were not just pulled out of the air. They were most likely sourced from press releases and other official documents intended for a similar purpose. Imagine the difficulties such choices could present to any bots that had to make investment decisions leveraged against Y2K spending on the basis of the above. In summary, the research illustrated that even trusted sources may have detectable editorial bias in their online offerings. This is something that is important to know and be able to measure the full extent of – an endeavour suitable for a further research project. It is also something that feeds directly into the questions of how a human or bot can be said to 'know' anything when such patterns dictate

---

[73] The repository can be found here:
https://www.dropbox.com/sh/qbufxoi0j85vik7/AACW7nvVWMqjBT0qPmGKgzY9a?dl=0

that particular narratives, which may or may not be true, enter the fray. How do we trust information delivered in this way if, even if the provenance is traceable (and the Cardiff research showed how often this was difficult at best), there are further editorial choices made that may obscure the wider picture and lead to multiple perspectives, each cherrypicking particular details, on the same event?

Reference, provenance and doubt meet computational coherentism

To restate my argument, the question of reference was considered following the narratives of bots in action was followed, making the trials of strength stand out and thus (hopefully) rendering the Black Boxes and circulation of reference visible (see chapters 3 and 4). A distinctly Latourian perspective was explored in order to both answer the question of reference and also to show how considerations regarding the practical description (and subsequent knowledge of) digital socio-technical systems themselves sidestep one prong of the philosophical sceptic's assault (the question of reference), making the latter somewhat irrelevant, at least in so far as pursuing the goal of obtaining "justified true belief" is concerned. As Latour, Whitehead and Harman show, the deployment of Propositions as a replacement for the fraught search for 'the true' is not only preferable – it is also much more reflective of the actual practices of individuals and institutions in seeking to justify knowledge claims. Combining the notion of Proposition with Circulating Reference yields a workable reliabilist account that can be applied to complex worknets. Any reliabilist account, by its very nature does not *guarantee* truth, rather it simply offers a higher likelihood of gleaning it using the method proposed. It is applied in this way as a heuristic method. Though it should be noted that reliabilist accounts still tend to regard 'truth' in a non-Propositional[74]

---

[74] In the Whiteheadian sense

way – i.e. they tend to regard it as singular and specific. The reliabilism at work in the application of Circulating Reference, bolstered with the idea of Whitehead's notion of Proposition is one of approximating the accuracy of the outcome as well as offering only a consistently likely (i.e. reliable) chance of this outcome. That is to say, it is entirely consistent with appealing to heuristics rather than methods that deliver certainties, absolutes and fixed entities.

Now I attempt to build on the broad argument and methodology sketched for apprehending reference as something mobile, here turning explicitly to the problem of competing lines of reference highlighted by the churnalism cases above. Understanding that reference, via Circulating Reference, is detached from a static state of affairs frees us from the ongoing difficulties in similarly static philosophical dialectics as to how claims of knowledge relate to their objects of study. The mobility also opens the door to having some chance at grasping dynamic worknets. It enables one to follow the narrative (the story) of a particular claim or set of claims about the putative object(s) of study. However, this leaves one precisely nowhere when there are competing narratives that attempt to occupy the same ground, even if only partially. They could be competing narratives regarding the same putative object(s), they could even be more or less identical conclusions reached with completely different referents or they could be a mixture of both as in the case of Pouchet's and Pasteur's disagreements - see Latour (1993a, 1993b, 1999). Why should one accept a single narrative over another if both are compatible with the facts as they are understood?

Whilst it is possible to conceive of interrogating each narrative in turn, then listing numerous reasons for and against with each position and settling (if possible) on what seems to be the most likely account, this is a potentially overwhelming feat when considering dynamic worknets. The speed and dynamism actively work against such an enterprise and continue the limited approach of effectively 'looking at events in the rear view mirror'. Even with the event

157

of the 2010 Flash Crash a number of years in the past at the time of writing, the untangling of narratives – and 'duelling algorithms' - is difficult (see Appendix 1). It is also less than useful if one of the aims here is to help address live sociotechnical systems and events that are unfolding even now. Even as you read this, the processes of churn detailed above will be in action, creating multiple narratives on the same events, people and even concepts that in turn will inform political and economic behaviour of individuals and groups.

How would this apply to the churnalism problems the chapter opened? A similar approach can be applied to bots where the provenance of a source is easily identified[75]. Sources of information can thereby be weighted accordingly if trying to weight the data a bot has gathered to form a coherent system. It may weight for example, data from the New York Times website higher than multiple pieces of data from a low profile blog. The blog data will not be immediately discounted, but would require not only multiple hits but also other parts of the system to cohere for it to displace the opposing datum from the New York Times website.

Interestingly, this also brings up the issue of using digital objects to provide reflexive feedback in a consistent way – effectively extending the capacities of interventions such as in (Rogers, 2006). Following Pollock and Thagard's thinking, one's assumptions and beliefs can be laid out in a way that has hitherto not been practical in a 'manual' context (i.e. *sans* computer). Thagard asserts that conscious human reasoning often follows a serial algorithm, leading to a somewhat "limited rationality". With the aid of computing technology however one can potentially engage in *conscious* coherence inferences (rather than serial inferences)[76],

---

[75] Usually a straightforward matter as a bot or human can easily visit the source of a transmission; outside hacks that deface or otherwise modify the data are usually quickly rectified.

[76] He distinguishes two views of inference – one clearly meant to represent the analytical mode as "a series of steps, each with a conclusion following from a set of premises" and another, his preferred choice, "as a largely unconscious process in which many pieces of information are combined in parallel into a coherent whole" (Thagard, 2000, p. 3). Justification is at the heart of philosophy, whether it is in epistemology, ethics or any other area. For Thagard, both the Rationalist (Plato, Descartes etc.) and Empiricist traditions (Locke, Hume etc.)

which are

> nonmonotonic in that maximizing coherence can lead to rejecting elements that were
>
> previously accepted. It is because limitations of attention and memory may lead
>
> people to adopt local, suboptimal methods for calculating coherence (Thagard, 2000,
>
> p. 30).

The possibility of offloading, or indeed enhancing, much of our cognitive load via computing

is an exciting (and sometimes maybe terrifying) prospect, not only in philosophical and

psychological terms but also in terms of productivity. Cognitive outsourcing as (Shirky,

2008) details it is already commonplace using computing and networked technologies. Just

consider, for example, how often you make the effort memorise something now that you

know is easily retrievable on Google. It was also something anticipated by (Wiener, 1954)

both as a widely occurring phenomenon in the future and a source of many new Black Boxes

to take account of. To take such work seriously however, especially in the context of future

human intellectual development, requires addressing the perceived 'unreality' of the virtual.

Critics of Thagard such as (Guarani, 2007) point to this as the most damaging case against

work such as Thagard's. The sense of 'unreality' where the virtual is concerned it refers to a

refusal to acknowledge computing technology as an essential tool in not only making human

patterns of reasoning explicit, but potentially being able to reason faster, if not necessarily

always better with the aid of a computing substrate. Guarani effectively argues that there is,

and will be, continuing institutional resistance to such a thing. This complements Latour's

specific notion of Institutionalisation and Institution in that such developments directly

---

have been failures as approaches for providing justification. There are no indubitable proofs to be found exclusively in either internal reflection or empirical enquiry. This is where coherence of some form steps in. We do not need indubitable foundations – this has been a harmful fiction imposed upon us by the analytic tradition – rather we justify beliefs, practices and principles through adjusting the latter until they reach a coherent state of relationships relative to one another. This contrasts with "the neat Euclidean picture of foundational axioms yielding a set of fully justified axioms" (Thagard, 2000, p. 5). Moreover, rather than a vague sense of coherence, what we seek, for Thagard, both consciously and in conscious justification is a *maximally coherent* set.

challenge the practices of the pre-existing Institutions that are used to maintain knowledge, which of course generates resistance. Similarly, just such a conflict is traced out neatly in several of Latour's examples such as the debate between Pouchet and Pasteur. The latter could also be modelled in coherence terms effectively with the use of negative 'constraints' that would prevent new information or activity from perturbing an existing coherent theory, narrative or point of view. What it means to include 'constraints' however, requires further explanation.

Choosing between elastic narratives

Once one has established the elastic narratives in play and transcribed them to a Thagard-like model, one has a fast assessment tool in hand for selecting from several apparently coherent or consistent narratives, where changes in any one element or assumption can easily be swapped in or out and the outcome recalculated. This might not certainly be 'pre-hoc' or anticipatory, however it is certainly no longer laboriously post-hoc. Answers to questions can at least be acquired in this combination of methodologies relatively rapidly. One can address the problems outlined in the opening sections of this chapter by both being able to detect possible churnalism and also choose between competing narratives (which may or may not be elastic in the sense discussed here).

An immediate problem presents itself, however. What if the output produces (within a certain range), equally coherent outputs? What if the stories about a bus being on the moon from the Daily Star are equally as coherent as those from the Independent that dismiss any vehicle other than a few Moon Landers and rovers being on the moon? Doesn't this return us to square one? It is important to remember that we are applying here a *heuristic* and as such, should not characterise it in short term and immediate gratification terms. The price to pay is that over the longer term (which in computing terms could actually just be minutes or hours)

our heuristic approaches will deliver reliably, however there will be gaps. Furthermore, even in its own terms, the computational coherence approach necessitates self-correction and as such *takes a longer view of multiply coherent systems coming into conflict with one another*. Moreover, some of those multiple systems may concern the viability (in coherentist terms) of one of the single elements in the main coherence problem we wish to solve - which of the two press outlets is more likely to be giving us accurate information about what is on the moon.

Thagard adopts this kind of 'long form' coherentism of which Bonjour is suspicious but I think has significant merit. It is an obviously possible outcome that two competing coherence problems could emerge as equally coherent, (or so close as to be negligibly different). In which case there is the difficulty of how to choose between them. Thagard bites the bullet and agrees that an immediate resolution may not be possible. However, on his account, with privileged information / input playing an important role and non-monotonic reasoning being a feature, he argues that the coherence is never settled in stone. It is an ongoing competition and assessment that will eventually force the issue one way or the other. Moreover, the theories would not just compete against of eachother in isolation, they would also have to accept (e.g.) empirical input. As Thagard notes in the contrast between Astronomy and Astrology, Astronomy has the weight of empirical evidence on its side and therefore over a period of time should expect to provide the most coherent outcome. Similarly, the 'no lunar bus' narrative postulated above is backed by widespread observation of the side of the Moon facing the earth, and less regular imaging of the dark side by occasional visiting satellites. To have a chance of maintaining empirical coherence and consistency, the 'lunar bus' narrative would likely have to assert the bus is situated on the less observed side.

Apposite to this is the fact that the coherence problems contain negative constraints such as inconsistency and as such this should also work in tandem with the greater empirical input for

Astronomy to ensure that it is resolved as the most coherent system. Not only that, but Astrology must also compete with many other theories with significant empirical confirmation. As Thagard argues

> Astrology might be taken to be coherent on its own, but once it offers explanations that compete with psychology and astronomy, it becomes a strong candidate for rejection. The isolation objection may be a problem for underspecified coherence theories that lack discrimination and negative constraints, but it is easily overcome by the constraint-satisfaction approach (Thagard, 2000, p. 73).

In addition to this kind of observation, it is noteworthy that sometimes inconsistency between systems is a feature, not a bug of scientific and epistemological enquiry. Sometimes mutually inconsistent systems have to be temporarily accepted as this also maximises (temporarily) coherence. The most famous contemporary example of such dualism is Relativity and Quantum Theory. Both are tremendously consistent and powerful explanatory approaches in their own right and yet are currently incommensurable (except via the distinctly *un*empirical string theory).

Summary

Can computational epistemologies in general and Computational Coherence in particular help one keep up with multiple elastic narratives and empower one to make choices between them? Can coherence, even supercharged with computing power enable an effective accounting of a Flash Crash or Ohio cascading power outage event, or a choice between narratives of dubious, possibly churned, provenance? I think it can, though not on its own. A perfectly legitimate line of criticism that has historically been leveled at coherence theories is

to point out a kind of hand-waving vagueness with regard to what coherence itself means. Especially with regard to how it is articulated and understood and how, practically, one goes about assessing it. Foundationalists such as Bonjour have often taken this line as it contrasts sharply with the careful axiomatic arguments that can be found in foundationalist accounts. It has force as an objection despite the fact that coherentism holds, as Bonjour admits, the greater intuitive appeal as it likely more closely reflects the processes of human reasoning and judgement.

In contrast to pre-computing accounts however, Thagard's is exceedingly sharply defined. The computational element and specification of modes, principles and algorithms largely rebuts the vagueness critcism. True, in some parts it is possible to argue that Thagard has essentially just moved a problem rather than solved it (such as, for example, how different types of coherence problem, or mode can interact computationally). I think whilst it is a fair criticism it does not detract from the overall success of these contributions in significantly progressing the project of practical coherentist approaches. In line with the overall philosophical mindset guiding this thesis, Thagard's approach works best in tandem with complementary approaches such as ANT, Circulating Reference or even Floridi's Philosophy of Information and Information Ethics. Any and all of the former are able to fill the 'specification' gap in Thagard's work. That gap is in determining what x affects which y and stipulating what values by which they should be judged. Regarding the latter, Floridi would provide the directives of IE or Latour would offer up the 'co-definition' actors and actants construct of one another and the controversies, ontologies and so on that go along with those specific accounts.

I would not agree with Thagard's assessment that, "Computing coherence can…be as exact as deduction or probabilistic reasoning" (Thagard, 2000, p. 70), except in the broadest sense that a very clear methodology (and input of appropriate elastic narratives) allows its

163

application. Those 'moved problems' that Bonjour highlights still haunt it, albeit not intractably. Many rapidly advancing computing fields such as Natural Language Processing and Machine Learning mean that the gritty work of specifying some of the trickier elements (such as semantic similarities, constraints etc.) are now within the grasp of any philosopher with access to a modestly powerful home computer. Indeed, Thagard concedes exactly this

> Being able to do this computation does not, of course, help with the problem of generating elements and constraints, but it does show how to make a judgment of coherence with the elements and constraints on hand (Thagard, 2000, p. 70).

In terms of its application, it is important to remember that this is a process that aims to provide *justification*, not *truth*. The concern regarding the lack of privileged information is definitely a sticky problem for any who assert a coherentist theory of *truth*. Less so for theories of *justification* in which the criteria for truth can be more widely construed and seen to include privileged inputs such as sense data or more trusted media sources. Coherentist theories of truth would make the very notion of coherence the guardian of truth (in a more absolute sense) and would thus allow for completely coherent systems of belief that bore no relation to the external world whatsoever. This is clearly unacceptable, which is why Thagard opts for models that explicitly privilege empirical 'real-world inputs over algorithmic models that simply seek to provide mathematical optimisation and with no direct link to privileged contributions or an avoidance of non-monotonic assessment. The connectionist approach by contrast explicitly includes privileged elements and gives them a particular strength in the network as well as maintaining the essential non-monotonic nature.

Moreover, given we have only ever been concerned here with epistemological *justification*, not truth, it provides an excellent heuristic. Such an approach seeking reliable justification combined with non-montonicism preserves the dynamism provided by Circulating Reference in creating elastic narratives. The dynamic *relationships* in flux that, through their

description, enable us to assess competing narratives and allow anticipation of scenarios such as transmission of churned content. If attempting to achieve the loftier goal of truth, we would be back to the tedious situation of dealing with petrified, static accounts. And the only way such accounts can encompass the dynamic sociotechnical world we wish to observe and describe is through being superseded by another equally petrified and static account. Although static accounts are not always bad, they are necessarily backwards looking. Indeed, I explore a static account of Object Oriented Programming in the following chapter. Insisting on a *truth* yielding rather than *justification* yielding mechanism in this context, however, would be like going backwards from motion picture to laboriously painted watercolour stills. We can reasonably obtain (limited) truths through static accounts (e.g. examination of source code). Epistemological justification on the other hand, in fast-moving environments where the code is being actively executed require heuristics that can only deliver justification.

Another essential dimension added through the application of Computational Coherentism is the capacity to cope with the dimension of *speed*. Moreover, through modifying the relative strength or weakness of excitatory or inhibitory links it can also take account of interacting worknets that are operating at different speeds. In describing the events of the 2010 Flash Crash for example, one could assume that the HFTs (High Frequency Trading bots or algorithms) are responding to the latest information whilst the poor human traders are still somewhat in the past, so any assessments originating from the latter should be regarded as discoherent or inhibitory. Though of course, as the discussion of competing narratives in Appendix 1 implies, these techniques could in fact end up just becoming part of the arms race between bot and human traders. The human assessments, without computerised aid in the form of something like Computational Coherence would appear discoherent or inhibitory *relative* to the HFTs. This does of course assume that one knows they are operating at

different speeds. However, such information can be established by working through the pattern(s) of Circulating Reference at work.

If exactitude is the goal then there are serious difficulties here, however as a heuristic approach that tends towards the optimum answer, then it's quite an impressive toolkit and could be of significant assistance in both apprehending and even to a mild degree anticipating Flash Crash and Ohio power outage type occurrences and enabling us to sort wheat from chaff in whatever competing narratives we are assessing, whether they are general media stories, formal accounts of events or – importantly – competing interpretations and narratives whilst the situation is ongoing. There is still some clarificatory work to be done, however. In the next chapter I make some comparisons with the practice of Object Oriented Programming attempting to reintegrate what has been discussed thus far as not only a collection of mutually useful methods but also as part of a wider general approach that assist in tackling our numerous digital apprehension problems.

Chapter 6: Case Study: Object Oriented Programming

I argued in the previous two chapters that elastic methods based on a computing substrate are necessary to effectively observe and describe similarly elastic worknets and events that are manifested and deployed partially or primarily via computing technologies. There are however further insights to be gleaned from this kind of active symmetry. Object Oriented Programming (OOP) has particular characteristics that provide an idealised and abstracted encapsulation of the ongoing 'negotiation' between human and machine in talking about and interfacing with the world in both digital and non-digital contexts in ways that were anticipated by Latour.

This chapter applies some Latourian concepts to Object Oriented Programming, providing some additional insight into apprehending digital objects in sociotechnical systems. The virtual objects of Object Oriented Programming provide exceptional clarity for the issues raised here. They do *not* represent in any way the foundational 'base level' or 'base concepts' required or sought after by foundationalist epistemologies. What they are is *articulation tools, par excellence*[77].

An Object Oriented Programming tool is a Black Box easily re-opened, at least when it is provided in an open source format. For example, the Object Oriented Programming tools provided by the Python programming language represent the congealed labour of absent actors who took part in developing the open source language, including its creator, Guido van Rossum[78]. Object Oriented Programming is almost a reification of the insight that there are

---

[77] This applies equally well to procedural programming in that the latter will make use of modules/subroutines in a similar way to Object Oriented Programming. Everything argued here applies as much to procedural programming, however Object Oriented Programming is given the focus because it explicitly articulates abstract relationships that are also highly portable.

[78] An interesting detail here is Guido's stated primary aim for creating Python. It was an adaptation in response to observing how programmers actually behaved. He introduced the controversial practice of internally defining

'absent makers' made present everywhere for both humans and non-humans and through both synthetic and natural means.

Putting the 'Programming' into a Digitised Latour

Latour does express a notion of 'program', though it is very particular. The *associations*, in whatever human/non-human configuration that we find, Latour also describes as intelligible *as* 'programs' (complete with 'anti-programs'). These programs go through processes of association/alliance on the one hand and 'substitution' on the other. They directly affect, or can even become part of, the specific association (actor-network) under consideration at the time. It is at this point that Latour particularly clearly expresses a philosophy that is not only 'Object Oriented' – it is also eminently suited to digital objects (associations) of study, as highlighted in this diagram from *Pandora's Hope*:
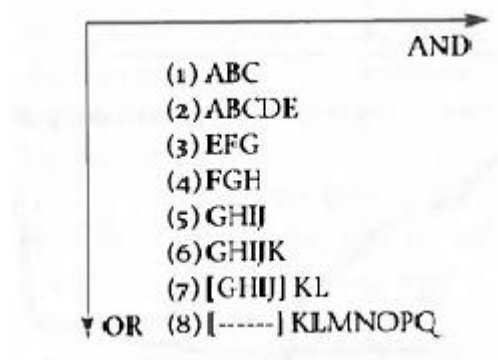

FIGURE 4

'PROGRAMS' IN PANDORA'S HOPE, (Latour, 1999, p. 160).

The above neatly demonstrates the process Latour envisions: As more associations are made, additional 'letters' (programs/objects/associations) are added to the 'word' – moving in the horizontal plane. However, for associations to continue, up until the point of Institutionalisation, substitutions must be made, so numerous letters change in the sequence.

decision structures such as loops with whitespace. This was because, he argued, human programmers spend more time looking at code than writing it.

At some point at least part of the 'word' becomes Black Boxed in an institution. This is represented by the bracketed sequence at step 7) and then followed by said Black Box becoming 'invisible' from step 8) onwards and hence regarded as a stable entity in its own right (where they "become matters of fact, and then matters of course").

The Blackboxed sequence [GHIJ] is said at this point to, at least in relative terms, 'exist'. Latour refers to such sequences as a 'syntagm'. There is thus a clear empirical process that can be followed without having to either assert an 'essence' or establish some kind of abyss crossing correspondence. It also allows for a considerable range of variations in understanding any said actor-network or syntagm. The strength of the claim to 'reality' or 'existence' of the entity referred to in the syntagm is a result of the length of the associations and the stability of the connections at work.

Far from courting the dangers of an extreme range of relativism, Latour argues instead that such an approach furnishes us with a:

> *Collective history* that allows us to judge the relative existence of a phenomenon; there is no higher court that would be *above* the collective and *beyond* the reach of history, although much of philosophy was devised to invent just such a court (Latour, 1999, p. 162).

We can, Latour argues, identify clearly the *envelope* of associations and substitutions that make up our studied worknet in the form of such a Black Boxed sequence – the 'program'. Each worknet would have its own unique signature in this regard, with specific associations and substitutions recorded and traceable. A comparable idea to heuristic approaches used, for example, to track and identify malware by anti-virus vendors. With regard to the different theories held by Pasteur and Pouchet, Latour argues that it is not necessarily the case that Pasteur and Pouchet are even targeting the same object of study, the same worknet, and

coming to different conclusions. They are dealing with, by definition, *different* worknets. Both have observable spatiotemporal envelopes that are sharp. The two positions are however incommensurable and Pasteur's wins out in the trials of strength. They may even as Latour notes, have begun in much closer agreement, however as the passage of time (and the experiments of Pasteur and Pouchet) continued, the series of associations and substitutions unique to each approach meant that the two assemblies had fewer and fewer elements in common.

Without an underlying law of inertia to drive it, it is down to Latour's notion of Institutionalisation to maintain the Pasteurian phenomenon in existence. And, to reiterate again, *work* must be carried out for this to be so. Further, it is something that can be verified by most of us wherever we are right now

> In whose world are we now living, that of Pasteur or that of Pouchet? I don't know about you, but for my part, I live inside the Pasteurian network, every time I eat pasteurized yogurt, drink pasteurized milk, or swallow antibiotics. In other words, to account for even a long-lasting victory, one does not have to grant extrahistoricity to a research program as if it would suddenly, at some threshold or turning point, need *no* further upkeep. What was an event must remain a continuing event. One simply has to go on historicizing and localizing the network and finding who and what make up its descendants (Latour, 1999, p. 168).

This emphasises once more how theory often fails to capture practice and the laser focus and straight lines of analytic philosophy are of limited use in contexts where we must follow the labyrinthine paths of machines, scientific practices and similarly rhizome like 'hairy' (as opposed to 'bald') "facts".

Every time a new link is made between two actors/actants, the two are somehow modified by the event. An example Latour uses to illustrate this is the gun control debate. He points out that the two main positions taken in the debate are unnecessarily polarised between giving complete autonomy to either the human or the machine (the gun). The debate necessarily becomes less polarised if we consider that as a human, we extend the possibilities of the gun, and as a gun, it extends the possibilities of the human. In short, "you are a different person with the gun in your hand" (Latour, 1999, p. 179), albeit as Latour is at pains to point out, not necessarily a different person in a *moral* sense. The same principle applies to the cognitive abilities of humans enhanced by computing technologies.

The better approach, Latour argues, is to approach both entities in question as Propositions. Regarding their mutual encounter as an event, he argues, emphasises that neither are fixed. And when the two Propositions encounter one another they are articulated into a *new* Proposition. It is still possible to articulate the attribution of a prime mover, however a composition of forces is required to explain any action. As actants may exchange competencies, or be socialised into a collective in the case of fabrication, so do they similarly offer one another new possibilities and functions in use. Object Oriented Programs similarly enlist one another as objects, spawn new objects and share (or divert) properties and methods. Unless the code is closed source, they are also Black Boxes that are readily opened by simple inspection (though "multiple inheritance" of properties can be tricky to trace back to the originating object or class or objects).

These mediating processes and roles are difficult to observe in sociotechnical systems, particularly in their earliest stages, because of Black Boxing. Often the very aspect we wish to measure is made opaque by Black Boxing and with it any insight into the joint production (or action) of actors and artifacts. Actions thus become properties of the whole association, the Black Boxing of which (or Black Boxing of components of) is made all the more

confusing by the fact that the number of perceived actants itself can change from one stage to the next. When something is Black Boxed it generally becomes 'invisible' and an actor/actant in its own right in any worknet mapped. The choice at which particular level of granularity we seek to rest our analysis is usually an arbitrary choice (or one limited by practical constraints). Any time we choose to begin opening Black Boxes in earnest, it is turtles (Black Boxes) all the way down:

> How far *back* in time, *away* in space, should we retrace our steps to follow all those silent entities that contribute peacefully to your reading this chapter at your desk? Return each of these entities to step 1; imagine the time when each was disinterested and going its own way, without being bent, enrolled, enlisted, mobilized, folded in any of the others' plots. From which forest should we take our wood? In which quarry should we let the stones quietly rest? (Latour,1999, p. 185).

If, however we are looking at Object Oriented Programs, they provide a *temporarily* articulated base level. That is to say, they provide very clear start and end points. The object references to other objects and digital entities is a chain that only goes a finite distance, even in extremely complex code. The difficulties in observing and apprehending such code's behaviour in practice is a factor of the networked dynamism in play as programs encounter one another and multifarious outputs in broadly unpredictable way. Looking at the code statically, however, still enable us to glean some important insights and a base from which to work.

The first object to be created, or called in the program, or used as a container for the others offers a convenient starting point – at least *within* the program itself. It is also, of course, arbitrary. The program in question could have been executed by another and operates within a context of many other programs, networks and sociotechnical worknet. What is important is

172

that this is implicitly recognised within the *practice* of programming in this way. For example, the core routine used to create a virtual programming object in the first place is called the *constructor*.

The desire for a foundationalist resting point (and the root of the regression problem detailed in Chapter 2) is rooted in the simple but necessary finitude of inquiry. We want to find something that can be written off as 'base reality' – a notion that may be hopelessly married to an infinite regress when considering digital systems in a sociotechnical context (Bostrum, 2003). Latour bites this particular bullet and regards it as nothing to be inherently concerned about. It is a problem that affects foundationalist accounts[79] no less than coherentist accounts. The advantage of the latter over the former being that it only ever seeks fuzzy boundaries and rough approximation where the epistemologies based upon the former persistently seek straight lines that may never be found. As Latour puts it:

> The depth of our ignorance about techniques [technologies] is unfathomable. We are not even able to count their number, nor can we tell whether they exist as objects or as assemblies or as so many sequences of skilled actions. Yet there remain philosophers who believe there are such things as abject objects (Latour, 1999, p. 185).

Language is also at stake here in that, by its very nature, it provides a foundationalist and petrified (i.e. static) means of communication, necessarily bending the meaning of the sentiments it carries in that respect. Coherence theories, their applications and how they fit into difficult yet subtle problems such as those considered in this thesis are not particularly amenable to linear exposition. As Latour notes in *Pandora's Hope* with regard to the use of technologies (techniques) by the researchers he accompanied into the Amazon, the supposedly commonsense boundary between 'signs' and 'signified' becomes rapidly

---

[79] Compare for example, the verification problem in programming (Fetzer, 1988).

confused. Our techniques change not only the language we use, but also the extent of its very functionality. A view that becomes all the more apparent if a fundamental abyss between words and world is not asserted and a theory of Circulating Reference, or something similarly dynamic, is used in its place.

Attempts to stabilise meaning and some kind of Circulating Reference can be clearly seen in Object Oriented *Programming* – the deployment of (virtual) "objects" and "classes" with "methods", "attributes" and "inheritance" follow a concrete pattern; with each stage assumed as a Black Box. Such practices give what Latour would call *meaningful articulation* and cross the sign/signified rubicon continually:

> What they can do is to explore the associations and the substitutions that trace a unique trajectory through the collective. Thus *we remain in meaning but no longer in discourse;* yet we do not reside among mere objects (Latour, 1999, p. 187).

The above could be straightforwardly applied to the 'collective' of Black Boxes that is a computer operating system. The example becomes much more interesting however when one considers a peculiarity of the networking infrastructure computer engineers originally chose to form the backbone protocols of the fledgling internet: that *the channels of communication are simultaneously the channels of control* (Zittrain, 2009). It is a *prima facie* example of the confusion of sign/signified, particularly with regard to the deployment of language that Latour seeks to highlight in so many of his empirical studies[80]. That in this case it applies primarily to computer (and hence non-human) actor-networks makes no difference. Our actual use and deployment of the notion of 'objects' is similarly made very clear in the *practice* of Object Oriented Programming. I contend that it often provides ideal use-case

---

[80] It also represents a popular idea in mythology and fiction made very real – the uttering of words simultaneously communicating and exerting power over the world – such as myths surround the control of demons through knowing their 'true names'. One can exert similar levels of control over machines with the correct incantations of *protocol*, *username* and *password*.

examples of the very practices that are much 'messier' beyond the digital world amongst the imbroglios of scientific research in 'meatworld'. The assertion of an 'object' with sharp lines is possible in Object Oriented Programming in a way that it is not outside the synthetic digital realm and where it is used as something more fundamental and enduring than a Black Box (i.e. when used to re-generate the subject/object words/world abysses), it is forgotten that it is an assertion and a temporary stopping point because

> An object *stands in* for an actor and creates an asymmetry between absent makers and occasional users….In delegation it is not, as in fiction, that I am here and elsewhere, that I am myself and someone else, but that an action, long past, of an actor, long disappeared, is still active here, today, on me. I live in the midst of technical *delegates;* I am folded into nonhumans (Latour, 1999, p. 189).

The properties of physical objects, such that they can be determined, appear in the context of tests, not as properties in and of themselves in isolation. Realism drives Latour's ontological conception because physical objects are able to "push back". If one thinks of human and non-human actors as equivalent, it is easy to see what Latour means when he describes physical objects of study as having "interests". Researchers in the 'hard' sciences attempt to isolate and manipulate the 'interests' of these objects, however they resist that manipulation and push back against the worknet trying to manipulate them (of which the scientist is part) (Latour 1993; Latour and Woolgar 1986).

Programmers, by contrast, create objects that push back in a completely artificial environment that is nevertheless hooked into the physical world and able to exert effects and similarly, be affected by the external non-virtual world. The Ohio power outage is a perfect example, where a feedback loop between the two generated a catastrophic physical failure.

For Latour, when sociology turned its gaze to 'hard' science, it encountered participants who would and could push back (in much similar fashion to their physical objects of study) in a way that participants in social science studies would or could not. Or were prevented from doing so (usually by a sociological interpretation that 'explained away' objections). For ANT, the 'hard' sciences have become powerful because their translations tend to be rigid (unlike the social sciences) and alternative translations are easily detected and/or amended where necessary. This is further reinforced by the process of Institutionalisation, where work is carried out within the institutional worknets to maintain these rigid 'translations' (Latour, 1999). This concept is used as a means to explain how, in a Latourian conception, sociotechnical objects or worknets can have any kind of inertia. OOP objects are completely rigid and as such are able to exert tremendous influence when leveraged properly.

Latour also focuses on what one might call the 'materiality of relations'. It neatly creates research questions that can be straightforwardly (if not without significant effort) deployed via analysing the components and links within any given worknet. Where the social sciences are concerned, Latour says that his aim is to "redefine the notion of social by going back to its original meaning and making it able to trace connections again" (Latour 2005, p. 1). It is no longer possible, he argues, to distinguish the connections when anything 'social' is asserted – there is too much assumed action at a distance and too many 'Black Boxes', such that it is not possible to construct a 'worknet' to explicate whichever social theory is being used. A "social context" has become a "default part of our mental software" – and in fact so-called 'social explanations' have become 'common sense' in this regard. Specific social ties have been taken to indicate the hidden presence of 'social forces'. ANT theorists, Latour argues, reject this assumed social context and explanation as action at a distance that *does not* account for the work it does, thus re-emphasising the empirical focus. Similarly, it is a mistake to adopt any kind of action-at-a-distance account of virtual objects. Their being

understood in a semi-chaotic networked environment is frustrated by the limitations of human cognition and even the limits for computers modelling one another's behaviour (to put it crudely, to model *n* computers for example you would need at least *n+1* computers). OOP objects are, in the Latourian sense, 'fully deployed'. Every possible interaction with them is spelled out clearly. Thus, investigating OOP source code before it is executed and released in 'the wild' provides a very neat 'materiality of relations' for understanding at least in controlled scenarios where inputs and interactions are limited, exactly how the code may play out in interactions. This is precisely what is done in 'unit testing' where code is exposed to a controlled series of inputs in order to check that it is behaving appropriately.

Code is created and deployed in social environments. Latour turns the conventional notion of 'social' on its head. It is not the 'universal glue' used to stick together and fill the gaps between all of the other disciplines and their own worknets and agents, nor a "spooky" concept existing over and above its constituent parts (Stirner, 1995)[81]. On the contrary 'the social' should be regarded as *what is glued together* by the connectors from all of the other subjects – it is the aggregate of these things, not some hidden force behind them over determining their own natures and behaviours. The active task of 'tracing the social' is instead for Latour, the tracing of *associations*. Crucially these associations are made via non-social means, ranging from inscriptions on paper through to world spanning communications networks. Code is the latest, and potentially most pervasive kind of inscription in human history to create, trace and describe these associations.

To emphasise again, science and technology, for Latour, do indeed face something very real. His philosophy is neither idealism nor postmodernism. Rather, he is shifting focus to what

---

[81] "Spooks" is a term used by Stirner in his sustained critique of collectivist terms that are used in a sense that places them on a metaphysical plane above the individuals that make them up. He says for example, regarding the 'Humanists' of his time that (note the emphasis) because they "served *Man*, they cut off the heads of *men*." (Stirner, 1995, p.103).

practitioners of science and technology *actually do*: they socialise (in the sense described above) nonhumans in such a way that they can meaningfully and observably bear upon human relations. This has been the history of technology/technique in perpetuity and as Latour points out, we would have to go very far back in history to find a point where this no long applies. Indeed, we would need to go back to prehistory to find a moment when humans (who we would likely have trouble even identifying as such relative to our current self-conception of 'human') were not fundamentally intertwined with techniques/technology. Even making allusions to primates would be insufficient as they too were never without technology (e.g. communication and social behaviours that enhance survivability), despite an assumed fundamental break in primitivism between human and primate, deploy techniques. Though an interesting point that Latour makes several times regarding the example of primates is that, lacking enduring techniques[82], they have to carry out continual *work* to maintain their social systems and hierarchies (Latour, 1993b, 1999). Our *object oriented* techniques on the other hand enable the maintenance of these associations with much less effort through physical artifacts and symbol systems.

What's in a word?

In this 'packing' of the world into words (and the many other means of articulation STS and ANT scholars identify), researchers are able to make the world mobile, manipulable and apprehendable. Especially for other actors distanced from the primary phenomenon under discussion: "Instead of moving around the objects, scientists make the objects move around them" (Latour, 1999, p. 101). The very tools one uses to grasp the dynamic world must, themselves, be dynamic. The common attempt to, by contrast, permanently fix petrified words to a similarly petrified (and hence necessarily *false*, in Latourian or Whiteheadian

---

[82] e.g. such as symbol systems, inscriptions or as Alfred Korzybski calls it 'time binding' (Korzybski, 1995) capacities.

terms) object of reference is the cliff upon which many philosophical endeavours have been smashed. Far too much weight is pressed upon the shoulders of the subject (in there) and the object (out there) to account, in static terms, for their truthful relationship – if there is one. Yet both are only two actors in an infinitely expandable worknet; the weight instead should be felt by the reliability of any proposed worknet connecting the two – as in Circulating Reference. This also presents something eminently empirical to investigate and account for – for example in Latour's notions of a 'spatio-temporal envelope' or Institutionalisation.

Similarly, the flow of inquiry appears incomprehensible if content and context are fundamentally separated ("enucleated" to use Latour's turn of phrase); a practice that of course makes complete sense when creating inscriptions to communicate but commits the category mistake, again, of mistaking the tool used to make the world mobile (via Circulating Reference) with the world itself. Abstraction is, like a Black Box, and like the work of Object Oriented Programming, something temporary and always to some degree constructed and artificial[83]. It is used to make communications portable and becomes a source of thorny problems if used and regarded as an end in itself. Such separations were/are for convenience and communication, not reflective of wider rifts in the actual nature of reality and lived experience.

An important further condition applies here. Latour's overall approach collapses content and context, intermixing them in contrast to what Latour refers to as "Model 1 science" – "conceived as a core surrounded by a corona of social contexts that are irrelevant to the definition of science; thus internalist and externalist explanations have little in common" (Latour, 1999, p. 92). The fundamental (which, in practice should always have been *temporary*) stopping points in our epistemological chains, or actor-networks, are not necessarily base elements of reality but Black Boxes that could be opened at any time,

---

[83] Creating the 'immutable mobiles' that Latour occasionally refers to.

179

revealing further networks. That "reality" pushes back is in no doubt, whether we can be said to grasp the whole through the interaction with the part that pushes back however, is in doubt[84].

How does one deal with such 'messy' epistemologies and ontologies in practice? Latour's solution is disarmingly straightforward: One does not have to abandon ship, rejecting all of the previous classifications for phenomena that had to be "purified" into one camp or the other (nature/science, subject/object, content/context), instead one can begin at either "end" and work towards the middle

> Instead of two histories which do not intersect at any point, we now have people who tell two symmetrical stories which include the *same* elements and the *same* actors, but in the *opposite order*…..Instead of cutting the Gordian knot – on the one hand pure science, on the other pure politics – it struggles to follow the gestures of those who tie it tighter (Latour, 1999, p. 86-7).

This echoes Thagard's 'long form' Coherentism, where multiple passes over coherence problems, each approaching from different angles at first should converge on overall inter-model coherence. Latour is at pains to point out that there is no *a priori* link to be assumed between content and context or 'science' and 'society' and that to follow the data honestly one may just as well discover that there is *no connection* at all, at least within the timeframe and or circumscribed worknet one is attempting to trace. The only *a priori* step taken at all is to refuse to make a fundamental division (purification) between the actors one may be following, or to put it another way "the distinction between internalist and externalist explanations is exactly as small (or as large) as the chain of translation itself" (Latour, 1999, p. 93).

---

[84] And it is this issue that motivates Harman's Speculative Realism.

Following this reasoning, if one were tracing an actor-network historically, the specific terms used at the beginning may well differ from those at the end. This once again reflects the dynamism inherent in his approach: actors, networks and actor-networks go through *translations* in the process of the continuous shifting of links, tests of strength and the movement of Circulating Reference. This is why one may begin talking about 'pure' scientific terms and end up using 'pure' political terms in their place. An extremely common transformation in the many actor-networks reflected in scientific controversies and practices that (Latour 1999; Woolgar 1991; Law 2004) and many others have endeavoured to describe in Actor-Network terms, and (Thagard, 2000) has characterised in coherentist terms.

Digital structures often carry with them the means to determine their provenance and capabilities – XML for example is a widely used standard for creating 'self-describing' data. If seeking a strong epistemological commitment to *know* what a digital entity *is*, one must simultaneously be able to apprehend its structure, whether it is a program, a network, or most commonly, both.

No wonder we have difficulty divining how to describe complex and fast moving assemblages – descriptions in common currency of our *methods* of description are themselves idealised and not reflective of actual practice. How such a view explicitly addresses questions of truth, *contra* Floridi's conception of it[85], Latour asserts: "There do not exist true statements that correspond to a state of affairs and false statements that do not, but only continuous or interrupted reference," *pace* Circulating Reference this has important implications for assessing any statements made by an individual. Or indeed inscriptions carried by any non-human (digital) entity, such as Object Oriented code, if one dispenses with concerns about intentionality, as Latour and other ANT practitioners do. The words do not matter so much as

---

[85] Floridi deploys a 'Correctness Theory of Truth (CTT)' with regard to Semantic information (Floridi, 2011).

how the person articulating a position is connected to the object(s) being referenced. The more connections there are and the more work they do in the worknet, the more "real" the articulated proposition(s). It is, as Latour puts it, the contrast between the 'highly connected' and 'sparsely connected', where the latter are "limited only to words" (Latour, 1999 p. 97).

One can in fact make meaningful claims to knowledge about the world even though the basis for such claims is indeed constructed (Latour, 1993b; 2005). There is no fixed list of ingredients that are combined in new ways. It is not a zero sum game. It is important to realise, Latour asserts, that a scientific experiment is an *event*. Unexpected elements almost always come into play and something very new may be produced at the other end, including entities that were not necessarily present at the beginning of the experiment. This is why metaphors play an important role in Latour's consideration of how to stabilise and describe entities.

Propositions by contrast enable us to capture the complexity and multiple agents at work, and explicitly force us not to hide them away, or stand in some staid proxy to represent them, complete with a footnote stating 'will deal with the noisy little monsters later…' The meeting of OOP objects in both static source code and executed programs provide just such Propositions in the Whiteheadian sense. In applying Propositions as one of the tools of choice one no longer asks anyone to utter statements where there is an inexplicable jump from words to world. Whether it ever has the precise contours of a mathematical axiom however is another matter (Whitehead may think so as he was as much a mathematician as a philosopher). That is the case whether in terms of usefully describing the situations referred to or in developing tools to describe, capture and assess them. Truth and falsity is now replaced with the *well-articulated* and the *inarticulate* Propositions. Heuristics developed on this basis won't have the sharp precision of Absolute Truth but then shouldn't we be asking

whether and *why* we would even want this, at least most of the time, other than 'Kant told us so'?

Summary

The new capabilities offered by computing technologies, themselves non-humans enrolled to human purposes and into human collectives – are what have allowed the enormously long chains of translation, reference and mediation that define, for Latour, the 'modern' in any meaningful way. There is – as Latour argues carefully in *We Have Never Been Modern* (Latour, 1993b) no fundamental break between 'primitive' and 'modern' societies; there is only – at best – a quantitative *not* qualitative difference:

> The difference, rather, is that the latter ["modern"/"advanced"] translates, crosses over, enrolls, and mobilizes more elements which are more intimately connected, with a more finely woven social fabric, than the former ["ancient"/"primitive"] does. The relation between the scale of collectives and the number of nonhumans enlisted in their midst is crucial. One finds, of course, longer chains of action in "modern" collectives, a greater number of nonhumans (machines, automatons, devices) associated with one another, but one must not overlook the *size* of markets, the *number* of people in their orbits, the *amplitude* of the mobilization: more objects, yes, but many more subjects as well (Latour, 1993b. p. 193-4).

Relative to earlier societies (and a number of currently existing societies still in our collective global midst) this quantitative difference is *huge*. With the introduction of digital and networking technologies, the jump is positively *astronomical*. Thus leading to the primary concern of computer ethics, that still rings with novelty, regarding *what machines (non-humans) do on their own:*

> They have been socialized to such an extent that they now have the capacity to create an assembly of their own, an automaton, checking and surveying, pushing and triggering other automata, as if with full autonomy (Latour, 1993b, p. 206).

The activity of distinguishing the above is of course entirely practical on a somewhat discrete and arbitrary basis – that is through a careful empirical description. The boundaries and Black Boxes will always be somewhat fluid and vague. Latour simply asks that we, as theorists, recognise and embrace this in the way that practitioners already do (whether they are consciously aware of doing so or not). And with the help of some of the tools, concepts and methods proposed in this thesis, such an enterprise can be further extended to the fast moving and digital too.

If an Actor-Network theorist were to observe the practice of Object Oriented Programmers and how they went about solving ICT problems using these programming techniques, they would note that the *practice* in question implicitly recognises the messy relationship between words and world. They, almost intuitively, deploy multiple epistemologies and ontologies as the situation demands. The objects of Object Oriented Programming not only share eachother's properties and spawn new instances, the humans do this too socially through the sharing and reuse of code. Human programmers essentially share *articulations* with one another via code that articulate sociotechnical realities or events with varying levels of success. The 'tiny gaps' of subject/object and measuring instrument/measured are crossed and re-crossed continually by these programmers. One runs into great collective difficulties however when attempting to describe what the programmers are doing and anticipating problems such as the 2010 Flash Crash or 2003 Ohio power outage because our use of language in these contexts still originates in the numerous fundamental divides that Latour eschews. It is no wonder that the application of both mathematics and computing technologies to the world in so many dramatic manifestations is seen as baffling to many

people because we still collectively labour under a fundamental linguistic and conceptual split between subject/object whether we realise it most of the time or not and whether one inserts such metaphysical assumptions into one's code can potentially make a great difference to its behaviour, though it seems that code as it exists in the virtual environment often forces a glossing over of such a distinction in direct virtue of being virtual itself, regardless of what assumptions the programmer(s) may carry.

A lot of ground has been covered so far in this and the preceding chapters and much of it quite abstract. I now tie it all together with some concluding thoughts in the final chapter.

Chapter 7: Conclusion

The kind of Object Oriented *Philosophy* raised here aims at a better kind of analytical clarity, not a multiplication of subjectivity. The classical distinction of subject/object created a sharp edged clarity that was based on a "polemical tug of war". Sharp edges are nevertheless still to be found in the alternative approaches proposed here, just located and expressed (articulated) differently. "We", both humans and non-humans and the multitude of hybrids exist in collectives. Empirically such collectives can be observed, differentiated, traced. Digital objects of study can provide particularly clear cut synthetic elements that are much more amenable to a Latourian treatment and thus traceable and accountable in these ways.

The modernist point of view on science and technology, as Latour sees it, is that they (science and technology) follow a continually progressive path of separating subjective and objective, subject and object ever further apart. Latour (at least in his primary texts considered here), is not aiming to dilute empirical inquiries or capabilities and neither am I. Quite the contrary:

> The name of the game is not to extend subjectivity to things, to treat humans like objects, to take machines for social actors, but *to avoid using* the subject-object distinction *at all* in order to talk about the folding of humans and nonhumans. What the new picture seeks to capture are the moves by which any given collective *extends* its social fabric to *other* entities (Latour, 1999, p. 193-4).

I have followed a similar path in situating analysis of networked assemblages within a framework that includes Circulating Reference and Computational Coherence. A particularly tantalising prospect offered here is that of being able to construct epistemological principles, ad hoc in digital spaces, that are perfectly suited to the particulars of the environment(s). For example, one would want to apply different principles in the cases of being limited to either Facebook or Myspace as they are built and function quite differently. In quite concrete terms

they have different distinct structures and ontologies. In the case of Facebook, one's epistemic warrant for believing that the entity one is interacting with is a stable identity (and even that biographical details are correct) is relatively justified thus giving a substantial epistemic warrant for further beliefs and inferences. Whilst in using Myspace one's reason for believing this would be much weaker because MySpace does not attempt to enforce the provenance of identity as Facebook does. MySpace encourages fanciful profiles whilst Facebook attempts to enforce 'real' profiles through means such as insisting on 'real' names. Myspace users have a lot of flexibility in designing the "experience" visitors have upon browsing their profile. Facebook by contrast provides a relatively generic one-size-fits-all template. One's 'epistemic algorithms' in either case then would look very different and yet be precisely suited to the contexts in hand. Circulating Reference could help us trace such differing provenance and Computational Coherence could similarly help us choose between them.

Such dynamic approaches would be appropriate in cases of adapting to conflicting signals and accounts regarding what is, or what just has happened in situations with fast networked assemblages at the centre such as the 2003 Ohio power outage. Once it has become apparent that the reliability of a source has been compromised or appears incoherent relative to others, one's decision making algorithms can take this into account. Moreover, every step of this can be performed in a computing substrate using automation. The way I carried out the churnalism research is a case in point – automation made it possible to very quickly assess the impact of press releases in the national press from the two sources considered. A consistent pattern was identifiable. One's assessment of both epistemological warrant and ontological provenance can then shift very quickly. Bots that have to respond quickly to new information or changes in external circumstances (such as in trading systems) behave in precisely this fashion. Our understanding of them routinely goes awry because we stubbornly continue to describe and model them in static subject/object and completely separable

epistemology and ontology terms. In the lightning fast world of digital communications, it no longer makes sense to hold on to the old approach of attempting to fix identity in stone and claim something definitively is A or B. Too many entities defy such categorisations faster than a poor human can apply them. Computing powered heuristic approaches that have a flexible or rather - *elastic* approach however, have better potential to grasp what is occurring. Not perfectly, but then nowhere has the cliché of 'the perfect of the enemy of the good' been more appropriate. And we can make significant progress beyond Circulating Reference and Propositions by using a computing substrate. If this is conceded, then what possibilities open up for Computational Coherence if used routinely as a tool for analysis?

Some form of Circulating Reference provides an invaluable heuristic even at early stages in sketching the 'vague boundaries', or 'envelope' as Latour calls it, of sociotechnical imbroglios. Anti-virus software works on a similar principle – viruses are identified by a signature profile; having particular properties, behaviours and algorithms. Whilst such an approach can often lead to false positives, and polymorphic viruses[86] present a particular challenge, it is still a proven methodology even if one does have to constantly update the virus database to take account of new signatures. This does however mean steering well away from the 'safety' of absolute certitude here; heuristics may not just be the best method for getting approximately right answers, they may in fact be one's *only* option. With the increasing importance of Big Data issues and cloud computing infrastructures such an approach is becoming the norm as algorithmic approaches such as MapReduce and services that track Twitter trends provide answers that are *good enough* rather than certain. Humans are fundamentally intertwined at every level of a networked assemblage, even when they are not directly present – programmers, policymakers and users specific epistemologies and ontologies, not to mention mistakes, accidents and ommissions, live on in the software

---

[86] Viruses that can adopt multiple forms and often reprogram themselves.

and code at work. Latour observed such phenomena at work in the laboratory and empirical gathering of evidence. And I have attempted to extend that work here in looking at the digital and networked domain, combining an updated Circulating Reference with Computational Coherence.

Excluding humans and the 'social' from such an analysis has been a mistake, as I hope I have made clear throughout. For Latour, with regard to experiments, the experimenter also adds something of themselves, in spite of the prohibition on such interference; though not in any mendacious way but rather by filling "the gap between underdetermined facts and what should be visible" (Latour, 1999, p. 128). Similarly, for the computable and the digital. Large data sets allow for a potentially infinite variety of experiments especially when combined with other linked data. When one is testing for *correlation* for example, one is also testing for *coherence*. Hypotheses are put forward and in some very direct sense, the scientist nudges their object(s) of study forward. The more metaphors and filters that one deploys, the better and with reference to the usual desultory references to 'rose tinted glasses' and similar optical metaphors for expressing bias:

> Far from opposing filters to an unmediated gaze, it is as if the more filters there were the clearer the gaze was, a contradiction that the venerable optical metaphors cannot sustain without breaking down (Latour, 1999, p. 137).

Using the language of 'the social' is appropriate at times – as a shorthand (for example using collective terms such as 'IBM', 'Britain', 'Social Capital'), permitting of course that one remembers it *is* a shorthand for something far more complex[87]. It is also more appropriate for organisations or states-of-affairs where events and changes occur very slowly and rarely.

---

[87] There is effectively a relation between collective nouns, shorthand, Black Boxing, and reification.

With regard to the study of the fast-moving objects of study for science and technology studies however, it is entirely inappropriate. Latour puts it thus

> in situations where innovations proliferate, where group boundaries are uncertain, when the range of entities to be taken into account fluctuates, the sociology of the social is no longer able to trace actors' new associations. At this point, the last thing to do would be to limit in advance the shape, size, heterogeneity, and combination of associations. To the convenient shorthand of the social, one has to substitute the painful and costly longhand of its associations (Latour 2010, p. 11).

The above sentiment applies especially to the massively more heterogeneous and mobile systems of networked assemblages. One of the key reasons for adapting Latourian concepts for these newer innovations and subjects is because they rapidly "multiply entities" and applying the 'common sense' notions of the social to these, "one then has an absolutist framework generating data that becomes hopelessly messed up" (Latour, 2005, p. 12).

In Latour's *philosophical* approach – all entities are regarded as contiguous and relevant and the main differing factor is the depth or distance of one's microscopic or macroscopic focus respectively: Yes, everything from atoms (and smaller!) through to macro social connections and up to celestial bodies are considered fair game. All are conceived as representing networks (worknets) at different levels and none are irrelevant. The choice of where to designate the cut-off point is almost always arbitrary, if ultimately necessary for any kind of progress and is characterised above in the nouns and verbs of Black Boxes and Black Boxing. This is a particularly suitable approach for a domain where so much is virtual.

Why is this relevant to the 'hard' (or 'harder') sciences and in particular to the problem of apprehending networked assemblages presented here? Because even if the scientific practice in question does not assert superfluous entities or relationships (unlikely), the worknet within

which it is situated (e.g. the laboratory, the computer, the network, the process) *does*. And many such aspects of the worknet of which the scientist or programmer is part will be informed by the social (including, naturally, the political). In the case of all of the entities involved in the examples of sociotechnical imbroglios considered so far, one is not just considering a complex sociotechnical system but a sociotechnical system that sits within, or adjoins many other sociotechnical systems.

In attempting to trace such complexity, the methodology for applying Circulating Reference essentially requires one to "make displacement more costly and more painful" (Latour 2005, p.22). That is, to adopt techniques of observation that forces one to instinctively avoid assumptions of 'action at a distance'. The first way to do this is to 'deploy' what Latour refers to as the "five major uncertainties"[88].

One of the things the researcher must do, in Latour's view, is to leave any defining or ordering of 'the social' to the actors under observation. There is a mutual co-definition at work that is observable (Latour, 1993a). These definitions and orderings then become further data points for the person attempting to trace the relationships that exist. The key corollary to 'deploying uncertainties' here is also to identify controversies in the observed worknet. It is the existing controversies that tend to be the flashpoint and origin for all the associations and activities that one wants to observe. For example, with the first uncertainty Latour recommends 'deploying' – the very shifting nature of groups and who or what is in each group is a continuing source of 'controversy'.

---

[88] These are "- the nature of groups: there exist many contradictory ways for actors to be given an identity;
- the nature of actions: in each course of action a great variety of agents seem to barge in and displace the original goals;
- the nature of objects: the type of agencies participating in interaction seems to remain wide open;
- the nature of facts: the links of natural sciences with the rest of society seems to be the source of continuous disputes;
- and, finally, about the type of studies done under the label of a science of the social as it is never clear in which precise sense social sciences can be said to be empirical" (Latour 2005, p. 22).

For Latour, in other sociological studies there is a continual, unquestionable assumption of boundaries. Usually put there initially by the researcher. And this applies just as strongly to domains that specifically look at computing or networking issues. Deploying this first uncertainty means coming to the observation with a completely open mind as to how many groups there are, their composition and how they relate to one another. This is obviously especially important with regard to the digital. In particular, one should be prepared for contradictory group formations. The typical application of 'common sense' social science methods and interpretations would usually deny this possibility. Latour asserts that encountering such contradictory formations is actually the norm, not the exception. 'Common sense' approaches tend to impose an absolute and arbitrary frame of reference – the arbitrariness would not be a problem were it not also absolute. A more objective account is achieved by the ANT researcher however by becoming familiar with the ever shifting frames of reference that actually exist in the worknet(s) under observation.

A further point made by Latour in this regard is that already existing connections leave few traces behind for the researcher to pick up and so it is dangerous to presuppose them and then fit evidence into that presupposition. Such 'unquestionably existing' groups are taken by ANT scholars to be beyond the proper realm of study. Continually shifting connections however tend to leave a rich seam of traces to observe. Almost every one of us comes into contact with the continually shifting connections of the internet and digital communications technologies on a daily basis.

There is a lot of exciting work to be carried out yet in a world of computerised and mobile philosophies and sciences along these lines. I have argued here for a few ways to push thought into motion and motion into thought in such a way as to make our buzzing, beeping, digital brave new world that little bit clearer. And moreover it is a clarity that helps machines speak as much as it helps humans.

We have great difficulty in anticipating even commonly complex occurrences in digital technologies such as simple software failure, or network overload, because we have struggled, and still struggle, to describe it. Are there better ways for *how to apprehend* such dynamic fast moving worknets? I think there are, and deployment of the particular methods and concepts that have been discussed so far are exemplars of the kind of approaches that *could* work. As highlighted several times, though, they are not the only choices. Given the nature of the problems that have driven this inquiry, there are several key implications of the overall argument presented here that would guide one's choice in how to approach the overarching problem.

Regression to the Subject/Object

There is a concept in statistics known as 'regression to the mean'. It refers to the fact that many patterns can be observed where a kind of homeostasis is maintained and the phenomenon in question almost always moves back to its typical state (which may or may not be an actual statistical mean) following a perturbation. A similar pattern occurs in discussion of Black Boxes. When attentiveness lapses, there is a regression to terms and language that separates ('purifies') observed actors and actants into subject/object categories. Both language and practice reflect this – one talks about and acts *as if* numerous actors are subjects and others objects. This is often regardless of how detailed and sophisticated one's accounts of phenomena may be when looking at them closely. The very terms of reference typically guide us towards talking about singular and unchanging entities. Many of the entities we now interact with however are not quite so singular, nor unchanging.

It is not just the technologies themselves that have changed in typical day to day human experience since the beginning of the computing and information 'revolutions'. The *kinds* of relationships we as humans have with them have changed also. We have moved from

relatively static to highly dynamic relationships. Our tools and measures have yet to keep up properly. With the ubiquity of these dynamic techniques and objects, methodologies that recognise the hybrids at work including our concept(s) of human/technology are needed. In an era of rapid change, cybernetics and its inheritors (who include almost everyone cited in this work) deployed concepts such as Black Boxes to smooth over the jarring disconnects between the deployment of technologies and our full understanding of their implications and impact. That a fundamental part of that disconnect has been (at least in the 'modern' West), driven significantly by the artificial 'purification' of binary categories such as human/technology has been one of Latour's most useful insights in this regard. The Latourian notion of the 'folding' of artefacts and humans into one another also enables articulation of the exercise of influence by absent makers across time and distance through technologies and artefacts. Adapting Circulating Reference to the digital as I have done here, effectively updates Latour for the digital realm.

It is essential in order to avoid the continual regression to this artificial purification, whether subject/object, nature/science, theory/practice or any one of a litany of such divisions to use terms that constantly remind us of the mobility and dynamism at work. I have proposed that applying Latour's Circulating Reference to digital objects is an excellent means to do so. However, it is the 'circulation' rather than the reference that is important to be continually mindful of. The interplay of relationships, interests of actors and actor-networks and trials of strength never ceases. Even if we must pause for breath sometimes in order to write down what we have observed, it is important to use terms and methodologies that encapsulate movement when they are both written and read back. At an almost obtuse and meta-level such terms contain within themselves the movements they record and pass them onto the reader. Being reminded that a database is as much an ongoing negotiated *process* as a gathering of physical electrical and magnetic impluses, stabilises the meaning and dimensions

194

of databases in general, and *this* database in particular in the reader's, speaker's and writer's minds much more effectively.

Competing epistemologies and ontologies

Competing epistemologies and ontologies come to the fore continually in almost every respect, including in the definition of measuring instruments themselves. For example, even with the initial advent of computing, the capabilities of the computers determined what *could* be observed or measured. Making a fundamental separation between subject and object then led to a damaging mistake in the choices of instruments used. Because, as discussed in the introduction, it was the least computationally expensive statistical operation, a linear trend became one of the dominant norms in computer aided analysis – perhaps continuing the use of a stalwart descriptive (and in the case of regression, potentially predictive) method beyond its useful life in the age of 'small data' and equally modest calculating instruments of the time.

Because such practices may not have been explicitly or consciously articulated in these cases, many processes such as those found in nature would be erroneously reduced to linear patterns when in fact they followed complex cyclical patterns. Computing power has of course advanced to the point where such calculations and patterns can often be routinely used with only a modest additional computational cost. It is however a good illustration of how artificially separating the two (as per several other similar divisions Latour attempts to cross such as science/nature, technology/human and so on) has led to *less* clarity and not more. It is also an interesting counterfactual speculation to wonder how many advances linked to consistently identifying such patterns may have occurred earlier, even with modest computing power, if the separation of subject and object at work had been consciously expressed at the time.

Even with the reflexivity forced on us as researchers in using data that is highly dynamic and therefore never a singular well defined entity we tend not to spell out our epistemologies or ontologies that we work with, even in just operational terms. The surprise failure in Google Flu Trends analysis for example may be attributable to this, especially as it effectively used a brute force technique to notice the pattern of searches corresponding with later flu epidemiology. The brute force ontology of raw pattern matching necessarily dictated a corresponding brute force epistemology of correlation and vice versa, repeating the almost clichéd error of effectively confusing correlation with causation. The Google researchers were not of course claiming that web searches caused flu outbreaks and yet that model would have been indistinguishable ontologically from the way it was used predictively – i.e. as a proxy that said 'interpret this data *as if* such a causative connection holds'.

The approach that Mayer-Schonberger and Cukier identified, described in Chapter 1, between causation or correlation as a dominant mode comes back to haunt us again and reminds us that, in spite of this Big Data assumptive method defaulting to correlation, it remains a very specific epistemological and ontological position regarding the nature of what is 'out there' and how to observe or measure it. Philosophers are suffering the double ignominy here of on the one hand communities of practice carrying out philosophy when needed ad hoc[89] and on the other hand, being left to pick up the pieces when everything comes to a crashing halt. Thus leading to much navel gazing head scratching as to what could have gone wrong – whether it is with the Flu Trends predictive capability, the event of the 2010 Flash Crash or countless other "technology" related imbroglios.

One of the important aspects of competing (and unacknowledged) epistemologies and ontologies to note with regard to these scenarios is that various events, and relationships, played out at both human and machine speeds. Models of worknets that involve both human

---

[89] Instead of philosophy explicitly being acknowledged as a continual part of the process

and machine agents tend to primarily map only the one or the other, when the messy reality is that they are highly interpenetrated. Observational and epistemological strategies need to be equally hybridised if they are to keep up. As (Lee, Cheng and Koh, 2010) noted in their categorisations of market participants in the 2010 Flash Crash, numerous human-machine hybrids effectively changed gears when they changed trading strategies in response to market events.

It is not just our objects of study that are hybrids. Our best methods of studying them – to 'stabilise' them tend to be hybrids also. We only separate the methods into 'pure' types *after the fact*. To talk about our brave new sociotechnical world accurately we must be both constructivists and realists in our use of language in a way that recognises and expresses how we are both constructivists and realists in *practice*. The realism is at work in the constant attention to the *states-of-affairs* – the *Propositions* in Whitehead's terms, manifesting in front of us. That so many of these encounters occur beyond immediate human cognition, and nowhere so much as in the realms of digital interactions, takes heuristic approaches beyond the merely sufficient and into the necessary – the *de rigeur* even. With the accounts provided by Latour and Whitehead, "facts" are both heavily constructed yet made more "real" by ever greater numbers of links. Latour categorically does not deny some kind of "out there" reality pushing back; what he disputes is our grasp of it, which is always in flux. Hence the "construction" in anything labelled "fact", which must if we were to be philosophically sincere, be regarded as a temporary fixture and not a petrified fixture for all of eternity as some epistemological accounts are wont to do.

To maintain the motion of ongoing construction and realism, I proposed combining Latour's insights, particularly Circulating Reference, with Computational Coherentism. This allows for tremendous latitude in deploying and acknowledging multiple epistemologies and ontologies. It is by no means the only method and either component could be used with an

alternative concept of active referencing/provenance or mapping and weighing competing narratives, respectively. Moreover, a computational substrate (in computational epistemologies and Computational Coherence) is primarily proposed because what is being observed is partly or mostly played out via a computing substrate itself. This may not always necessarily be the case.

Using Computational Coherence to let the social flow back in

Mayer-Schonberger and Cukier's assertions regarding the amount of the data dictating whether curation and accuracy was possible don't unfortunately stretch to the other end of the scale, where there may be *too much* data and the currently established culture of Big Data (such as "letting the data speak") hinder as much as they help. Apophenia may actually be the last of our worries if we cannot tie our enormous collections of data and reserves of computing power back into a sociotechnical context. As the numerous narratives of the Flash Crash explicated in Appendix 1 indicate the human, social, contexts were often missing initially and yet crucial for explaining events[90]. There is *work* to be done in such situations that is often neglected.

Adopting a Computational Coherentist approach significantly mitigates Boyd and Crawford's concerns regarding apophenia with large data sets and also forces engagement with the social context. A limitation of Computational Coherentism approaches is that *someone* has to carry out the gritty detailed work of determining what all of the agents concerned may value. Yet this forces the outside, social, context to flow back in. Using both Latour's Circulating Reference as I have and his specific flavour of Actor-Network Theory methodology one could "deploy controversies" and "deploy uncertainties" in order to fill in these details (thus

---

[90] Such as the entirely human decisions made to engage circuit breakers or 'go slows' at different exchanges and at different times, creating staggered effects and shockwaves. It is also easy (ironic given the reminder in the text at this point of apophenia) to write these actions off as the result of just 'human panic'. These may have seemed perfectly rational, calculated actions to the actors at the time and their wisdom is only challenged when a much larger picture emerges later *post hoc*.

fulfilling the "specification gap" Thagard identified in first constructing Computational Coherentist models). Or if the epistemological inputs and outputs are of more concern (such as in the case of deciding whether one competing scientific theory is superior to another), the relevant relationships and truth bearing links could be established with Circulating Reference and then pushed into a Trial of Strength with one another via Computational Coherence.

Whilst it obviously does not remove the confirmation bias inherent in human pattern seeking and matching, especially as the specific operational statements that are compared to one another for coherence are articulated by humans, the *process* of continuous excitation of the coherence model is at least somewhat blind. Thus patterns of computed coherence and discoherence are at least likely to be more reflective and hence *descriptive* of reliable relationships and patterns in the data because they capture both dynamism at work in situ and built in contradictions that may not have been visible at first. Humans simply playing with the dataset looking for interesting gestalts that may or may not match their personal (and currently conscious) preconceptions is not necessarily the best methodology on its own (though it may certainly produce *interesting* results). Computational Coherence in this context also thereby provides a useful adjucation method between competing epistemologies and ontologies. For all of these reasons I argue that the computable approaches and the social approaches need one another. Each domain is potentially able to displace, even if just temporarily, the biases and limitations of the other. However, it is an ongoing process of reflexive feedback loops and competition between models of Circulating Reference to be more coherent (or discoherent) with others and certainly interesting for further research.

Reconsidering Institutions

"Costly and massive institutions" were of great utility for their time, the period of "the modern" (Latour, 1993b) that was pre-computing and pre fast moving networked

assemblages. They bridged the gap between the desire for absolute truth and the relative experience of reality pushing back. The institution provided *inertia* and this inertia is now thoroughly challenged by a world steeped in computing technology. Google, Facebook and Apple may all deal primarily or solely in the digital realm and as such are insitutions with inertia themselves. However, their current mode of adaptation is to usually absorb smaller competitors and upstarts, such as Facebook's recent acquisition of WhatsApp[91]. Something which isn't really adaptation at all and may in fact be simply putting off the inevitable in having to give way. Indeed, Facebook has yet to fully monetise its product and investment has proceeded on the basis of how valuable the enormous database of millions of human interactions would be to researchers. It is unknown territory where the very nature of an 'Institution' is shifting and in flux especially such 'hybrids' as Facebook and Google. A similar shift in understanding was required identified by (Latour, 1993a) when "Microbes" were introduced in the 19th century, and suddenly the number of entities multiplied into the millions. Now we grasp around for description in a world awash in *billions* of artificial agents.

Even prior to the computational turn, the 'useful relativity' of the Institution could only be maintained if we avoided the four "absolute adverbs" Latour identifies: never, nowhere, always and everywhere. They were already under strain in Pasteur's time (Latour, 1993a). Now they are positively at breaking point. These are artefacts of an absolutist – and in turn, subject/object, turn and the cause of their own problems. Such a rejection may exhibit fears invoking a dangerous relativism where truth or falsity cannot be qualified do so on the basis of a demarcation, (sharply and finally cut in dead inscriptions such as the written word), that does not pay the price of maintaining the institutions that in turn maintain facts durably. They

---

[91] See for example, 'WhatsApp: Facebook acquires messaging service in $19bn deal' https://www.theguardian.com/technology/2014/feb/19/facebook-buys-whatsapp-16bn-deal - last accessed 20 Oct, 2014.

assert a "cost-free inertia" and one not to be in either nature or society (found if we were to maintain the nature/society distinction). As Latour puts it, "*demarcation is the enemy of differentiation*", (Latour, 1999, p. 157, original emphasis). Worknets are differentiated, not demarcated as are the very scientific practices[92] that they aim to describe. Demarcation attempts to impose straight razor sharp lines where there are none.

Looking further afield then, there are tricky choices ahead to be made therefore with regard to the process of Institutionalisation as Latour described it. As Mayer-Schonberger and Cukier point out, whilst using the commonly fixed notion of 'Institution' rather than Latour's more process/practice oriented term, our Institutions of learning and research have historically been built on the curation of small (and often fixed) data sets. Data and understanding that is in continual motion along with its methods and instruments of observation and interrogation, raises significant questions regarding the still popular conception of an Institution as dwelling within a specific building or set of building. It may be the case that as cloud computing has completely shifted dynamic server provision to a virtual domain, where the physical configuration is only a bit player, the creation of MOOCs (Massive Online Open Courses) may prefigure future changes in our understanding of universities and learning institutions in general.

Concluding thoughts

The various technologies and approaches that are already being developed in Big Data clearly anticipate a number of the proposals and explorations considered in this thesis, even if these are not necessarily explicitly articulated as such in the Big Data literature. In (Mayer-Schonberger and Cukier, 2013) the argument that both Big Data and 'small data' are somehow factors of the instruments available at the time as well as raw volumes of data

---

[92] and the object(s) of those practices – entities that they attempt to stabilise.

speaks to the notion of using Circulating Reference to make the specific epistemology and ontology clear in making it clear and similarly making the reflexivity at work in all worknets more visible – there are no completely passive, detached observers as per the classical epistemological ideal.

It was noted in Chapter 1 that with the advent of non-relational databases, the ontologies and epistemologies at work could be altered with each change to the database. Relational databases could at least always be guaranteed to conform to their fixed schema. By contrast fundamentally different types of data can be processed side by side in NoSQL databases and no single ontology or epistemology is capable of capturing that, certainly not *a priori* in a specific sense even if a general ontological or epistemological account is visibly at work for the technologies as a whole.

All of the insights and concerns raised thus far directly speak to the questions of effectively apprehending fast moving networked assemblages raised in the first chapter. We do indeed have a difficulty – one that is, hopefully now well established here as a multifaced problem. All of them are traceable back to ontological and epistemological issues, even those of a psychological nature that aggravated Thagard and Pollock. Many Latourian concepts assist in bridging these gaps and making the description of one's imbroglios possible and intelligible once again. Inherently dynamic additions such as Circulating Reference also allow a kind of mobility to take root in our descriptions that more closely resembles the living, moving objects and subjects of study.

Accuracy, I agree, following Mayer-Schonberger and Cukier's reasoning, is indeed a luxury of small data sets and curation. My reasoning for agreeing is quite to different theirs, however. I would argue on the basis of the additional insights provided by Latour and Whitehead that this state of affairs is only the case because a smaller data set enables a more

complete enumeration of the sociotechnical relationships, actors and Propositions at work. Whilst aiming at accuracy is still an inherently valuable goal, it also become somewhat irrelevant in a world of mobile objects and datasets that change at genuinely lightning speed, thus rendering clearly the necessity of heuristic approaches. The perfect, as the cliché goes, is very much the enemy of the good in these cases. Circulating Reference as I have adapted it here provides an excellent heuristic to pursue at ponderous human speeds, however it soon lags behind as an effective technique on its own when integrating live action in computing substrates. Its use is not nullified; however, it becomes something more preparatory. A tool to sketch out the general picture of the dynamic relationships at once that is elastic enough to anticipate changes and differentials (such as in speed, distribution and so on) yet not sufficient for being of much use when computing speed enters the equation, except as a post hoc tool. It is from this point that I proposed computational approaches such as Thagard and Pollock's (and many others such as Floridi's) are able to fill the speed and complexity gaps with a fast moving heuristic that itself requires a computing substrate to even function. And once these heuristics are set in motion, description and knowledge, as competing ontologies and epistemologies become intertwined processes, only stopping if something breaks or for the purposes of abstraction and communication in a work such as this.

Sometimes it has made abstract sense to fundamentally separate humans and non-humans in our ontological and epistemological models in order to make a particular part of our shared reality stand out and be rendered visible. It is an unfortunate historical happenstance that such a separation persisted and was effectively petrified into disciplines and Institutions, handing down the separation for philosophers and scientists alike to continue. As (Zittrain, 2009), (Galloway, 2004) and (Lanier, 2013) observed just as the way the internet was physically and virtually structured could have been otherwise, the same applies to the historical accidents and maybe a few choices that resulted in our dominant subject-object paradigms of ontology

and epistemology that Latour and Whitehead challenged. They too could have been otherwise. It is just as unfortunate that as a result we also treated the human and the machine as entirely separate when, as Latour and many others have pointed out, we have never really been apart.

Appendix 1: 2010 Flash Crash Narratives

The Flash Crash – duelling accounts and reconstructions

On May 6th, 2010 something unprecedented happened to the U.S. stock markets. Within the space of ten minutes hundreds of billions in dollars were wiped off of the value of listed companies. Some companies were suddenly rendered worthless, priced at just one cent per share, only to recover suddenly to dollar double figures in seconds. Other stocks swung haphazardly in the other direction. The market recovered again within minutes – the rollercoaster went down and then back up again so quickly anyone who had taken so much as a tea break missed it.

The trades driving these events had been carried out faster than humanly possible by algorithms following the practice of 'High Frequency Trading' (HFT). These automated agents, or 'bots', would normally behave primarily in response to well understood general movements in the market and the investment strategies of their makers. In this case however they were, without human supervision, also responding to one another in spans of milliseconds. The event was subsequently dubbed 'The Flash Crash' and describing exactly how it was caused, or which algorithms were primarily responsible, is still controversial. An algorithm is at base a list of instructions that make decisions on the basis of contingent inputs, usually producing equally contingent outputs or triggering actions – in this case, instructions to buy or sell at certain prices.

User attention was present yet either incapable of grasping what was occurring or powerless to arrest it. The event(s) unfolded too fast, effectively making user oversight (and attempts at intervention) redundant. Those few user led interventions that did occur in a timely fashion, such as immediately pulling out of the market or the exchange imposing a 'go-slow' on

trading, only worsened matters. The high-frequency algorithms interpreted these as further signals to engage in a mass dumping of assets.

The differing accounts, both of what happened and who or what the prime movers were during the Flash Crash duel in a way that is reminiscent of the very same duelling that took place in those minutes of the Crash between bot and bot, human and human and bot and human. Numerous accounts sought to blame the bots (Barnes, 2010) – specifically the High Frequency Traders (HFTs), whilst others dismissed their role as secondary, identifying instead micro-structural features of not only the market where the problem started (the E-mini futures) but also those closely linked in equities such as the S&P 500 (Lee, Cheng and Koh, 2010).

Even after publication of both preliminary and final official reports of the CFTC-SEC (2010a, 2010b) confusion still reigned supreme. One interesting aspect detailed in the reports was that the traders responsible for the top 10 largest 'longs' and 'shorts' (essentially, trading positions that overvalued or undervalued) were on both sides of the market, which contravenes the dominant narrative of a one sided collapse of the market towards short selling. Whilst it is true that numerous prices were squeezed down to the ridiculously cheap, thus affecting everyone who held that stock or option, those affected most seriously *in situ* (i.e. as active market participants at the time) held ground on both sides of the market, indicating something had gone seriously awry in their trading strategy.

The preliminary report (CFTC-SEC, 2010a) proposed several hypotheses for the dominant causative explanation: dark pools (private trading and information matching networks) caused a sudden withdrawal of liquidity as they all trade on the fragmented market of U.S. trading venues. Alternatively, the circuit breaker at the NYSE entered a 'go slow' phase to allow liquidity to return, but orders were just rerouted elsewhere. Or perhaps the fact that two

exchanges were freed from their obligations to re-route orders because the NYSE was in 'go slow' mode and did not respond quickly enough. It could have been the fact that 'stub orders' (orders priced at ridiculously low prices in order to enable the market exist[93]) were executed when they never should have been, leading to a panicked pile on. And so on. All of these features played a role to be sure. The final report (CFTC-SEC, 2010b) concludes this issue with a crude chronological narrative beginning with the initial large trade ordered by Waddell & Reed, which simply diverts the inquiry into the psychologically comfortable terrain of a linear account regarding a multiply complex event that was anything but linear. The first trade (the 'large trader') is identified effectively as the first feature, and therefore the primary descriptive and explanatory touchstone also. This is despite the fact that several 'archaeological' reconstructions demonstrated that the same fund had carried out similar trades previously without the same result. The Flash Crash was in fact a perfect storm of many factors.

Take for example, the 'stop logic' functionality or 'circuit breakers' on numerous markets, including the E-mini futures. They were instituted in a staggered order thus resulting in even more chaos. The pause in the E-mini occurred at 2:45:28 for five seconds. However, the pause only meant that orders could not be concluded – it was still possible to enter and modify orders on the book. And algorithms did just that. One of the unusual consequences of this was breaking the relationship between the E-mini futures and the NYSE equity trading. The latter is supposed to drive the former, but events in the E-mini market reversed this. This also resulted, in turn, in the NYSE's own 'go slow'. (Lee, Cheng and Koh, 2010) argue that it is precisely this kind of 'changing the rules mid-game' that resulted in such perverse responses from pre-existing trading strategies, especially those that were automated. The consequences, they argue, go even further than this – for the subsequent cancellation ("bust")

---

[93] Known as 'stub orders', created by 'market makers' who often get some kind of benefit or preferential treatment for maintaining orders in the market.

of many trades (generally those that were 60% or more away from their prices prior to the crash) now creates concerns for market makers who previously would have stepped in to provide liquidity in similar situations now being unsure if their trades will be honoured (Steiner 2013; Patterson 2012).

(Lee, Cheng and Koh, 2010) assert that no model can simulate two particular features simultaneously –

i) the uneven rules of operation between trading venues (c.f. the staggered circuit breakers) and

ii) the dependency models (and actual behaviour) between different asset types (e.g. between equities and options/futures).

 It is crucially important to note difference between reconstruction and simulation here, lest we fall into the 'look ahead' bias[94] (Chan, 2013, p. 4). Reconstructions across the literature appear to be relatively well regarded – the worry is the inability to simulate and therefore predict another such event. Why this is the case is highlighted in an important feature of the reconstructions, such as (Lee, Cheng and Koh 2010; Cespa and Foucault 2014; Menkveld and Yueshen, 2013). The (Lee, Cheng and Koh, 2010) study concluded that mandatory position limits imposed beforehand would *not* prevent another Flash Crash. This is because, according to their reconstruction, volatility shocks occurred at the microsecond, not second, level – hence the circuit breakers may be worse than useless. Moreover, the various reconstructions tend to focus on only one of the affected markets. The data available from each market has its own specific *resolution* – i.e. the rate at which data was gathered and preserved (microseconds, seconds). Thus even an excellent reconstruction of events in one market

---

[94] A well known problem in attempting historical trading simulation, "using tomorrow's prices to determine today's trading signals" (Chan, 2013, p. 4). To be able to classify account holders at the time, and attempt to predict how they might influence the market would require future knowledge.

provides a description that operates at a pace that may not reflect developments in the linked markets. They also may only provide data on orders from the perspective of one participant or group of participants. Differential speeds, resolutions and perspectives thereby interject even more conflict between the various accounts.

Lots of tools and techniques have been refined in the finance industry to anticipate and integrate changes and internal shocks to the electronic market system. For example, (Chan, 2013, p. ix) details instruments for detecting reversion to the mean and cointegration of financial instruments. There is little wider awareness of the other worknets within which the markets sit however, including and especially the social relationships that are operative on, but sit outside of the market bubble. For example, an ANT perspective would have been more likely to note the speed differentials that exacerbated the Flash Crash symptoms and causes. Some of these are social relationships that effectively exist 'in the pipeline' - where the original decision making has an effect long after it was put into place. The 'circuit breakers' installed at the interlinked markets operate at 'human' speeds (i.e. within seconds), whilst the High Frequency Traders adapt their strategies in milliseconds. The 'stalling' of circuit breakers being switched on at different trading venues at varying times was also a human call.

As Chan notes himself, the rise of 'dark pools' has been "quite challenging" (Chan, 2013, p. x) for trading strategies. The 'dark pools' operate effectively like un-openable Black Boxes within the worknet, hiding potential secrets and surprises known only to those who are direct participants. (Patterson, 2012) asserts that their primary intention was to allow human traders to escape high-frequency robot traders engaging in hit and run tactics. However, their role soon diversified, and many such Dark Pools filled with bots just like their public versions. These dark markets could also be used by human traders to offer better prices to participants than those offered publicly.

These pools can frustrate regulatory attempts to control 'wild west' or 'casino' style trading by legal authorities such as the SEC and FSA. They also create large unknowns for any statistically based system of prediction and description that is not already an insider. Moreover, the dark pools can also operate at human or machine speeds or both at different times. In the former case they may be used for 'over the counter' trading and in the latter for high-frequency trading intended to work around awkward legislation or features of the current trading infrastructure (Patterson, 2012). Thus, there is a need to go beyond just adequate accounting of the sociotechnical features as suggested above and also use dynamic methods of description that allow for unknowns such as dark pools. This necessitates heuristic approaches wherever possible, so immediate accuracy is traded for elasticity in the shape of what is being, or *could* be, described.

Such heuristic approaches would also have to account for large scale tit for tat exchanges in the relationships being observed. The story is far more complex than merciless bots on one side and witless traders on the other. There are multiple sides and many 'flips' between them that can potentially occur between them for both human and bot participants. For example, in 2011, Credit Suisse Securities announced the launch of a new product - 'Light Pool'. It was designed by programmer, Daniel Mathisson explicitly to counteract the negative effects of both Dark Pools and High Frequency Traders. Akin to an advancement in a major arms race it deployed bots to detect other bots and exclude traders who behaved in a 'bot-like' manner in order to "set prices and keep out unwanted speculators" (Mehta, 2011). It is important to note that this did not discriminate against algorithmic trading, rather it aimed to defang high speed trading that was beyond trader (and investor) purview. A key aspect to this technology was that it would not distribute information about orders through an exclusive 'fast' feed that could be easily read and acted upon by bots. Instead, the information would go into a consolidated feed available to all traders and at a suitably slower speed. Mathisson is quoted

referring to these ultra-fast bots as the "pick-off artists". And in order to defeat them there would, tellingly, "be no Black Box." (Patterson, 2012). As Patterson put it,

> The Plumbers [network engineers, programmers] had always believed that a problem with the machine could be fixed with a better machine. But what if the problem wasn't inside the machine? What if it was the all-too-human arms race itself, a race that had gripped the market and launched it on an unstoppable and completely unpredictable path? (Patterson, 2012, Loc 188-91).

A different emphasis is provided by (Menkveld and Yueshen, 2013). They characterise the turn of events as a failure in the cointegration strategy followed by the original seller between the E-mini[95] and SPY[96] securities. The primary feature was a combination of human (pre)intervention through the circuit breakers (in this case, the temporary halt on trading on the E-mini market), followed by aggressive high frequency selling. Their paper uses data provided by the 'large seller' for a reconstruction. They also highlight that patterns of integration affected rippled out to other distant markets, as the IIORC report notes:

> It spread rapidly to trading in other index products, but also to individual stocks. The crash echoed internationally as, for example, Canadian markets also crashed after approximately two minutes (IIROC, 2010).

Menkveld and Yueshen dispute the 'single' cause or 'prime mover' theory supported by the official CFTC report, arguing that the crash could be the product of how agents interacted that day.

---

[95] Index futures based on the S&P 500.

[96] Abbreviation for "Spyder", a S&P 500 index tracker.

This is supported by other accounts such as (Madhaven, 2011) who also noted that the primary victims were outside the immediate envelope of the E-mini and SPY trades at the centre of the action: "Hardest hit were ETFs [Exchange Traded Funds] that traded in fragmented markets, most affected by high-frequency trading and intermarket sweep orders." Madhaven asserts that "dynamic interrelationships" were at work, rendering the single-cause argument untenable.

The official narrative claims the large seller (the first trader) could not find counterparties in the SPY market, this broke the "arbitrage" relationship between the E-mini and SPY markets, leading to aggressive behaviour by High Frequency Traders (including the original seller's) to close out positions as fast as possible in E-mini. High Frequency Traders have been observed to lead to these kind of "liquidity spirals" before (Brunnermeier and Pedersen, 2008)[97].

Following this key aspect of the event, (Cespa and Foucault, 2014) characterise the Flash Crash as "illiquidity contagion" that occurs as a result of "contagion through a learning channel". With regard to the narrative outlined by Menkveld, this refers to information being suddenly limited if one or more securities become illiquid. This has a knock on effect to other liquidity suppliers who limit the availability of their securities due to the reduced information pool. It becomes a feedback loop that inevitably leads to a crash.

This returns us to a key, possibly *the* key, dimension: Granularity of available data, and what was recorded as an "event" (or not) was critical in the squeeze of information between the traded commodities (E-mini and SPY):

> The E-mini contract trades only at the Chicago Mercantile Exchange (CME). The data
> feed is very detailed as each order book change or trade becomes an event and is
> recorded in the database. SPY, however, trades on eight different exchanges: BATS,

---

[97] Liquidity is the measure of how easy it is to sell or convert an asset, or move out of a holding position.

BOST, CBOE, CHIC, CINC, ISEX, NQES, and PACF. Its data feed is less detailed as only trades and changes in a market's best bid or ask are recorded as events (Menkveld and Yueshen, 2013, p. 7).

And between exchanges, "a depth change on the highest bid in BATS becomes an event, also when this bid is strictly lower than a bid from another exchange." To make matters worse on the informational front, long delays (in trading terms) are allowed, "exchanges are allowed a 90 second delay on reporting trades which shows in the data; more than 90% of the trades, however, seem to be reported with, at most, a 100 millisecond delay" (Menkveld and Yueshen, 2013, p. 7).

Menkveld's came closest to a 'gods' eye view' in terms of granularity – ticking at 25ms. These kind of differences in granularity are important to capture, and could be using a modified ANT-like methodology that described the general sociotechnical worknet that all of these actors and actants are part of. Describing these characteristics - granularity, speed differentials, circuit breakers, market links and so on still would not be sufficient however. They focus on the trading infrastructure. As important, if not more important, in painting an accurate picture is carrying out the same task for describing informational flows. As the official reports noted, even when a circuit breaker was activated, orders can still be entered, modified or cancelled even if they cannot be concluded. Actors would still be making decisions on this changing information even if they could not complete trades. In the case of HFT algorithms, they may well have been doing so in a Batesonesque "strange loop", making rapid fire recalculations and decisions over and other without a third party interjecting (whether a human or otherwise) with the important information that something had gone fundamentally wrong and thus such calculations and decisions were suddenly rendered pointless.

Moreover, the E-mini market experienced a halt (though not necessarily in informational trading terms), but the SPY did not, leading beyond just a staggered series of trades that had previously been cointegrated and into equally confusing and quite possibly contradictory information flows.

Much of this broken trade strategy and information integration was beyond human response times. Highlighting the level of precision at work here, (Menkveld and Yueshen, 2013) identify the break in cointegration in the 'large seller's' algorithms for the two markets at the following levels of precision

cointegration broke at 14:44:27.525 (1:00.625 before the halt) and recovered at 14:53:19.425 (7:46.325 after the halt) (Menkveld and Yueshen, 2013, p. 11).

These are, to human eyes, obscenely tight timescales. At this stage lots of epistemological quandaries were suddenly present that high-frequency algorithms were not equipped to assess. Nor were human traders able to make such assessments even when timescales slowed down to seconds. Both bot and human traders were expected to become, effectively, 'five second philosophers', figuring out what was real and what was not in an informational landscape that was changing at a blindingly fast rate. Regarding the opportunities presented suddenly in the market by the first trader's large sell order, Menkveld and Yueshen decided to solicit opinions of other participants regarding how their decisions may have been made:

Interactions with industry on this seemingly large arbitrage opportunity yielded several insights. First, given the enormous price decline and extremely low system response times, it was immediately clear that uncharted territory had been entered. Might orders execute on one leg of my arbitrage and not on the other? Perhaps they execute but get canceled ex-post due to extreme market circumstances (as happened for equity trades)? Second, could price series be trusted? At some point delivery of

214

the Dow Jones Index series was 80 seconds delayed. Third, margin requirements rose

rapidly on steep increases in volatility. Netting across markets was impossible for

most participants. Finally, the deal looked 'too good to be true' (there must be

something that others know and I do not know) (Menkveld and Yueshen, 2013, p. 11).

Menkveld and Yueshen also note that "timestamp synchronicity" plagues modelling and

reporting of these events given that records may be kept at different levels of granularity and

that "events" may be recorded differently between exchanges, markets and even traders. This

supports their narrative that characterises failure in a "disproportionate response" of other

traders in response to original sell order, rather than first mover trader. Attributing

responsibility is also difficult due to the speeds of the trades. All occurred at millisecond

timescales, so decisions were made by non-humans and were therefore out of immediate

human control. In a sense, even if the 'first trader' narrative held, their intentions did not

matter as the behaviour of all agents including their own algorithms were out of their control

as soon as they began executing. They, quite literally, engaged in thousands of Trials of

Strength with one another, with many actors/actants being circumvented and driven off

course by other market participants. It is important to note the parallel trial of strength for

most, if not all, participants also in process – the battle against an informational vacuum

regarding what was 'really' going on.

It was not possible to establish relationships between E-mini and SPY as they were in

perpetual flux, driven by the very actors that needed to know what they were and designed to

be adaptive to the behaviour of other participants. It was a perfect exemplar of chasing one's

tail. Menkveld and Yueshen contend that a further confounding factor was the model, which

would have been the *de facto* choice for most participants – the CRRA utility. They note that

it "does not admit negative wealth". Given the sharp swings up and down in stock values and

positions it would have been completely blind to "impossible" (if entirely temporary)

downswings in an actor's position in the market.

Menkveld and Yueshen conclude that

> A clear-cut recommendation does not emerge as it depends on the mechanism that
> drove the destructive interaction of agents in the market. Further study of candidate
> mechanisms is left for future theoretical work (Menkveld and Yueshen, 2013, p. 25–
> 6).

There is thus a wide gap in understanding remaining. Furthermore, Flash Crash type events,
at least on a smaller scale, have become commonplace according to the (Johnson et al., 2012)
paper *Financial Black Swans Driven by Ultrafast Machine Ecology*. The authors even use the
term 'sociotechnical' in attempting to describe the situation and future prospects of these
complex, highly networked and ultra-fast trading systems. They argue that in recent years the
markets have undergone "an abrupt system wide transition from a mixed human-machine
phase to a new all-machine phase characterized by frequent black swan events with ultrafast
durations.(<650ms for crashes, <950ms for spikes)." Moreover, the frequency is alarming:
They claim that there have been 18,520 "ultrafast black swan events" between 2006 and 2011
alone that are detectable in stock-price movements. They conclude that these events occur as
a direct result of operations occurring and speeds that are beyond human response times.

What is the possible end result of such unknowns, in such high frequencies? Patterson
considers the expectations of people he interviewed, involved the field, in the wake of the
Flash Crash:

> The Flash Crash had proven this wasn't merely a fanciful nightmare scenario bandied
> about by apocalyptic market Luddites. The question tormenting experts was how far
> the loop would go next time. Progress Software, a firm that tracks algorithmic trading,

predicted that a financial institution would lose one billion dollars or more in 2012

when a rogue algorithm went 'into an infinite loop … which cannot be shut down.'

And since the computer programs were now linked across markets— stock trades

were synced to currencies and commodities and futures and bonds—and since many

of the programs were very similar and were massively leveraged, the fear haunting the

minds of the Plumbers was that the entire system could snap like a brittle twig in a

matter of minutes. A chaotic butterfly effect could erase everyone's hard-earned

savings in an eyeblink and, for kicks, throw the global economy into yet another Wall

Street–spawned tailspin. The pieces were already in place. Exchanges from Singapore

to China to Europe to the United States were linking up through a vast web of algo

traders dabbling in every tradable security in the world. The threat had grown so

tangible that it even had a name: the Splash Crash (Patterson, 2012, Loc 189-207).

Others have chimed in on this kind of disturbing scenario. Professor Michael Kearns, an

architect of numerous Wall Street trading algorithms himself, doubts that cause and effect

narratives are even possible in the context of such speed and dynamism. He is quoted in

Wired, saying "Our financial markets have become a largely automated adaptive dynamical

system, with feedback.....There's no science I'm aware of that's up to the task of

understanding its potential implications" (Salmon and Stokes, 2010). The relationship

between machine and human however, is highly complex and nuanced. Moreover, it is a

complex interleaving and dance between the two that goes beyond just concerns about speed.

In what way is this true? Briefly considering part of the history of the ever increasing speed

of these interactions is enlightening - it suggests that humans adapt to each sea change in

speed in some way. In Latourian terms, each aggregate (assemblage) of actors finds ways to

subsequently subvert the other to its intended trajectory.

Peterffy and Jerecki in 1969 introduced one of the very first genuine algorithm based Black Boxes to the market "which inhale market data, chew on it, then issue an instruction to their user, in this case whether to buy or sell" (Steiner, 2013, Loc. 326-66). Moreover, as Steiner notes, because they had a much deeper understanding of mathematics and coding than their peers in the trading profession, Peterffy and Jarecki were able to implement cutting edge mathematical theories such as the Black-Scholes formula that was published in 1973. Automating the operation of this formula meant they effectively carried out cognitive outsourcing (Shirky, 2008) that their competitors could not and consistently made money as a result.

And yet, in 1977, Peterffy's Black Box was likely outmanoeuvred by humans. His algorithms failed when DuPont suddenly announced a stock split and he lost $100,000 of his capital. He claimed that people were trading with insider knowledge to which he - and by extension his algorithms - were not party. The social context re asserted itself and became something that Petterffy had to take on board in writing and deploying automated trades. This process of continual adaptation continued throughout the history of the U.S. stock markets. Ever increasing speed meant ever faster and higher volumes of trades taking place on a daily basis:

> All of that turnover was having a real-world impact on stocks. At the end of World War II, the average holding period for a stock was four years. By 2000, it was eight months. By 2008, it was two months. And by 2011 it was twenty-two seconds (Steiner, 2013, Loc 752-57).

Moreover, it did not matter how mathematically sophisticated modelling became, "No one was keeping their eyes on the Bots' activities. No one could, since no computer on earth could capture all of the manic nanosecond action" (Steiner, 2013, Loc 718-22). Worse, when

creating models there is a bias towards the mathematically simple and elegant – not taking account of the inherent messiness of human affairs and behaviour,

> It's easier to write algorithms to fit normal distributions. And despite history showing us repeatedly that human behavior is anything but normal, some hackers choose to account for only normal distribution. Using this assumption can make money 100 out of 100 days. But it's day 101, the Black Monday of 1987, the Russian debt default of 1998, the Flash Crash of 2010, that can ruin those banking on algorithms designed purely around Gaussian distributions. Even Gauss, more than two hundred years ago, warned that errors of any magnitude are possible within a normal distribution (Steiner, 2013, Loc. 1040-47).

It seems at these junctures, before there is a breakdown, technologists assume that their algorithms have got the measure of man. Conversely, wily traders and hackers in deploying insider knowledge or disruptive behaviours, assume they have got the measure of the machine. Neither are correct. As Latour puts it, "How many actants are there? This cannot be determined until they have been measured against each other" (Latour, 1993a, p. 164), and "Everything may be made to the measure of everything else" (Latour, 1993a, p. 158). As the regular occurrence of misunderstandings and minor disasters attest, something altogether more fluid is needed. Approaches that can take account of man and machine in a constantly moving, tested, stretched pattern of relationships. We can't just blame the algorithms, there is a social context that is ignored at peril. Hence the need for sociotechnical accounts.

Appendix 2: Robots.txt files

Unless overwise stated, files were accessed on 09/08/2013

Business category

British Chambers of Commerce
Original URL: http://www.britishchambers.org.uk/robots.txt

bcc.txt

User-agent: *

Crawl-delay: 90

Crawl-Delay: 90

BP
Original URL: http://www.bp.com/robots.txt

bp.txt

robots.txt for bp.com

User-agent: *

Disallow:

Disallow:/content/bp/en/esi/*

Disallow:/content/bp/en/ajax/*

Disallow:/content/dam/infrastructure/*

Disallow:/liveassets/bp_internet/globalbp/*

Confederation of British Industries
Original URL: http://news.cbi.org.uk/robots.txt

cbi.txt

robots.txt for Umbraco

User-agent: *

Disallow: /aspnet_client/

Disallow: /bin/

Disallow: /config/

Disallow: /css/

Disallow: /data/

Disallow: /install/

Disallow: /masterpages/

Disallow: /python/

Disallow: /scripts/

Disallow: /umbraco/

Disallow: /umbraco_client/

Disallow: /usercontrols/

Disallow: /xslt/

GlaxoSmithKline
Original URL: http://www.gsk.com/robots.txt

gsk.txt

User-Agent: *

Disallow: /investors/shareholder-information/corporate-sporsored-nominees*/

Disallow: /investors/shareholder-information/corporate-sporsored-nominees*.html$

Disallow: /*.htm$

Disallow: /*.js$

Disallow: /about/

Disallow: /about-us/our-mission-and-strategy/delivering-more-products-of-value.html$

Disallow: /about-us/our-mission-and-strategy/overview.html$

Disallow: /avandia/

Disallow: /careers/downloads/

Disallow: /careers/uk_downloads/

Disallow: /careers/gsk-apprenticeship-programme-2012.pdf$

Disallow: /careers/gsk-interview-handbook.pdf$

Disallow: /careers/talent/

Disallow: /collaborations/

Disallow: /community/

Disallow: /ControllerServlet?

Disallow: /downloads/

Disallow: /education/

Disallow: /en_GB/

Disallow: /global/

Disallow: /infocus/

Disallow: /investors/agm/

Disallow: /investors/downloads/

Disallow: /investors/our-year-in-review/

Disallow: /investors/presentations/

Disallow: /investors/product_pipeline/

Disallow: /investors/reports/

Disallow: /investors/reps*.pdf$

Disallow: /investors/shareholder-information/corporate-sporsored-nominees*.html$

Disallow: /investors/shareholder-information/corporate-sporsored-nominees*

Disallow: /media/05-appendices/

Disallow: /media/appendices/

Disallow: /media/consumer-advisories/

Disallow: /media/downloads/

Disallow: /media/flu/

Disallow: /media/media-contact-details/

Disallow: /media/london-2012

Disallow: /media/imagecatalog.htm?Pid

Disallow: /media/*.pdf$

Disallow: /media/paroxetine/*.pdf$

Disallow: /media/press-kits/

Disallow: /media/videos/

Disallow: /mission-strategy

Disallow: /products/vaccines.html$

Disallow: /products/prescription-medicines/

Disallow: /rare-diseases/

Disallow: /reports/

Disallow: /research/about/

Disallow: /research/clinical/

Disallow: /research/open-data.html$

Disallow: /research/open-data/

Disallow: /responsibility/cr_issues/

Disallow: /responsibility/cr_report_*/

Disallow: /responsibility/cr-review-*/

Disallow: /responsibility/downloads/

Disallow: /rss/gsk-features.xml$

Disallow: /ser/

Disallow: /uk/about/

Disallow: /uk/about-us/information-for-the-media/

Disallow: /uk/common/

Disallow: /uk/links/

Disallow: /uk/locations/

Sitemap: http://www.gsk.com/site.xml

Rio Tinto Group
Original URL: http://www.riotinto.com/robots.txt

rtg.txt

User-agent: *

Disallow: /video-library/

Disallow: /ourproducts/index_ourproducts.asp

Disallow: /rio-apps/riotinto_videorepository/

Disallow: /ourapproach/

Disallow: /library/

Disallow: /ourproducts/

Sabmiller
Original URL: http://www.sabmiller.com/robots.txt

smiller.txt

User-agent: *

Disallow: /files/reports/sdreporting/managed_content/library/

Vodafone (UK)
Original URL: http://www.vodafone.co.uk/robots.txt

vodafone.txt

Robots.txt file for http://www.vodafone.co.uk/

User-agent: *

Disallow: http://www.vodafone.co.uk/en_GB/assets

Disallow: http://www.vodafone.co.uk/en_GB/css

Disallow: http://www.vodafone.co.uk/en_GB/js

Sitemap: http://www.vodafone.co.uk/vodafone-uk/appsitemap/sitemap.xml

Media Category

BBC
Original URL: http://www.bbc.co.uk/robots.txt

bbc.txt

User-agent: *

Disallow: /cgi-bin

Disallow: /cgi-perl

Disallow: /lexaurus

Disallow: /mpapps

Disallow: /mpsearch

Disallow: /mtk

Disallow: /weatherbeta

Disallow: /weather/hi/about/newsid_7760000/7760846.stm

Disallow: /1/hi/help/8594343.stm

Disallow: /2/hi/help/8594343.stm

Disallow: /1/hi/help/8594488.stm

Disallow: /2/hi/help/8594488.stm

Sitemap: http://news.bbc.co.uk/weather/sitemap.xml

Sitemap: http://news.bbc.co.uk/weather/sitemap_mobile.xml

Daily Mail
Original URL: http://www.dailymail.co.uk/robots.txt

dailymail.txt

Robots.txt for http://www.dailymail.co.uk/ updated 04/12/12

TS-5682

All Robots

User-agent: *

Begin Article Rules

Disallow: /*/article-1234741/*

Disallow: /*/article-2278543/*

Disallow: /*/article-2142419/*

Disallow: /*/article-1268644/*

Disallow: /*/article-2153937/*

Disallow: /*/article-2278173/*

Disallow: /*/article-2256210/*

Disallow: /*/article-2256902/*

Disallow: /*/article-2256955/*

Disallow: /*/article-379930/*

Disallow: /*/article-2146565/*

Disallow: /*/article-1516173/*

Disallow: /*/article-1516756/*

Disallow: /*/article-1515083/*

Disallow: /*/article-1513374/*

Disallow: /*/article-1513815/*

Disallow: /*/article-1585241/*

Disallow: /*/article-1600925/*

Disallow: /*/article-1601836/*

Disallow: /*/article-1602668/*

Disallow: /*/article-1603610/*

Disallow: /*/article-1610647/*

Disallow: /*/article-1611105/*

Disallow: /*/article-1615201/*

Disallow: /*/article-1615663/*

Disallow: /*/article-1619040/*

Disallow: /*/article-1090536/*

Disallow: /*/article-1619863/*

Disallow: /*/article-1622239/*

Disallow: /*/article-1629909/*

Disallow: /*/article-1630829/*

Disallow: /*/article-1625247/*

Disallow: /*/article-1613830/*

Disallow: /*/article-1648361/*

Disallow: /*/article-1660124/*

Disallow: /*/article-1151728/*

Disallow: /*/article-1667485/*

Disallow: /*/article-1221098/*

Disallow: /*/article-1687150/*

Disallow: /*/article-1667104/*

Disallow: /*/article-1685002/*

Disallow: /*/article-1250815/*

Disallow: /*/article-1689307/*

Disallow: /*/article-1690401/*

Disallow: /*/article-1690463/*

Disallow: /*/article-1690944/*

Disallow: /*/article-1692185/*

Disallow: /*/article-1694150/*

Disallow: /*/article-1694526/*

Disallow: /*/article-1280488/*

Disallow: /*/article-1341576/*

Disallow: /*/article-2257948/*

Disallow: /*/article-2233054/*

Disallow: /*/article-2256939/*

Disallow: /*/article-2256939/*

Disallow: /*/article-2237655/*

Disallow: /*/article-2198515/*

Disallow: /*/article-2215080/*

Disallow: /*/article-1018904/*

Disallow: /*/article-1251554/*

Disallow: /*/article-1292332/*

Disallow: /*/article-1328592/*

Disallow: /*/article-566617/*

Disallow: /*/article-2156960/*

Disallow: /*/article-1202656/*

Disallow: /*/article-2227550/*

Disallow: /money/markets/article-2024243/Soci-t-G-n-rale.html

Begin Standard Rules

Disallow: *readcommentshtml*

Disallow: /home/christmas/*

Disallow: /home/scotland/*

Disallow: *columnist-/.html

Disallow: *edge.create*

Disallow: *jsessionid*

Disallow: *reportAbuse*

Disallow: /*?printingPage=true$

Disallow: /*?start=

Disallow: /*competitionId=*

Disallow: /*createThread.html

Disallow: /*debateSearchResults.html

Disallow: /*debateTagSearch.html

Disallow: /*debateUserSearch.html

Disallow: /*emailArticle.html$

Disallow: /*goto.php?

Disallow: /*login?redirectPath=

Disallow: /*logout?redirectPath=

Disallow: /*myStories.html$

Disallow: /*nextThread.html

Disallow: /*no-text*

Disallow: /*pageOffset

Disallow: /*pageSize

Disallow: /*previousThread.html

Disallow: /*questionId*

Disallow: /*readLater.html$

Disallow: /*redirect.php?

Disallow: /*refer_product.php?

Disallow: /*searchPhrase=*

Disallow: /*selectedImage*

Disallow: /*startIndex=

Disallow: /*textbased/channel

Disallow: /*threadIndex=*

Disallow: /*topGallery*

Disallow: /?article_id*

Disallow: /?in_article_id*

Disallow: /ce/item.cms*

Disallow: /dailymail/

Disallow: /debate/polls/poll.html$

Disallow: /destinations/

Disallow: /dwr/*

Disallow: /error/*

Disallow: /guide/*

Disallow: /home/search.html?explore*

Disallow: /home/sitemaparchive/static$

Disallow: /js

Disallow: /link/index.html*

Disallow: /mailshopQA/*

Disallow: /mobile/*

Disallow: /news/ireland/*

Disallow: /pages/*

Disallow: /reader-comments/*

Disallow: /registration/*

Disallow: /SITE=DM/

Disallow: /sport/columnist-1000543/hatchet-man.rss

Disallow: /sport/columnist-1031773/andy-roddick.rss

Disallow: /sport/columnist-1046072/glenn-hoddle.rss

Disallow: /sport/columnist-1060209/jake-humphrey.rss

Disallow: /sport/columnist-353/sir-bobby-robson.rss

Disallow: /sport/columnist-354/ian-ridley--mail-sunday-sports-reporter.rss

Disallow: /sport/columnist-1043711/mark-jeffreys.rss

Disallow: /sport/columnist-423/dan-king.rss

Disallow: /sport/columnist-352/monty-panesar.rss

Disallow: /sudoku*

Disallow: /travel/enjoyengland/index.rss

Disallow: /travel/visitcornwall/index.html

Disallow: /tvshowbiz/azstar/archives/calum-best.html*

Disallow: /tvshowbiz/azstar/archives/coleen-mccloughlin.html*

Disallow: /tvshowbiz/azstar/archives/gordon-ramsey.html*

Disallow: /tvshowbiz/azstar/archives/nicola-robert.html*

Disallow: /tvshowbiz/tvlistings/*

Disallow: /weather/*

Disallow: /ukplus/*

Disallow: */chat/*

Disallow: /cms/sites/all/modules/ckeditor_link/proxy.php*

Disallow Money

Disallow: /money/*

Disallow: /tmoney/*

Allow Adsense

User-agent: Mediapartners-Google

Disallow:

Disallow Specific Robots

User-agent: MnoGoSearch/*

Disallow: /

User-agent: omgilibot/0.3

Disallow: /

User-agent: WebVac

Disallow: /

User-agent: WebZip

Disallow: /

User-agent: psbot

Disallow: /

User-agent: daumoa

Disallow: /

User-agent: ia_archiver

Disallow: /*/article-*

Sitemap Files

Sitemap: http://www.dailymail.co.uk/newssitemap.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2013.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2012.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2011.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2010.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2009.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2008.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2007.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2006.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2005.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2004.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2003.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2002.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2001.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~2000.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~1999.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~1998.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~1997.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~1996.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~1995.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~1994.xml

Sitemap: http://www.dailymail.co.uk/sitemap-articles-year~1993.xml

Sitemap: http://www.dailymail.co.uk/videositemap.xml

The Express
Original URL: http://www.express.co.uk/robots.txt

express.txt

User-agent: *

Disallow: /myexpress/

Disallow: /printer/

Financial Times
Original URL: https://next.ft.com/robots.txt

financialtimes.txt

User-agent: Googlebot-Mobile

Disallow: /

User-agent: Googlebot-News

Disallow: /FTePaper

Disallow: /epaper

Disallow: /cms/s/8bb7fbd4-e176-11dd-afa0-0000779fd2ac.html

Disallow: /Common/

Disallow: /wrapper/ftvideo

Disallow: /alertshubwrapper11

Disallow: /thirdpartywarpper/managedfundshome

Disallow: /thirdpartywrapper/marketsdatahome

Disallow: /thirdpartywrapper/alertshubprefcenter

Disallow: /readitlaterwrappernorr

Disallow: /readitlaterwrapper11

Disallow: /alertshub-wrapper

Disallow: /mba-newslines-wrapper11

User-agent: Googlebot-Image

Disallow: /FTePaper

Disallow: /epaper

Disallow: /cms/s/8bb7fbd4-e176-11dd-afa0-0000779fd2ac.html

Disallow: /Common/

Disallow: /wrapper/ftvideo

Disallow: /alertshubwrapper11

Disallow: /thirdpartywarpper/managedfundshome

Disallow: /thirdpartywrapper/marketsdatahome

Disallow: /thirdpartywrapper/alertshubprefcenter

Disallow: /readitlaterwrappernorr

234

Disallow: /readitlaterwrapper11

Disallow: /alertshub-wrapper

Disallow: /mba-newslines-wrapper11

User-agent: Googlebot

Disallow: /search/

Disallow: /ftArticle

Disallow: /FTePaper

Disallow: /epaper

Disallow: /cms/s/8bb7fbd4-e176-11dd-afa0-0000779fd2ac.html

Disallow: /Common/

Disallow: /wrapper/ftvideo

Disallow: /alertshubwrapper11

Disallow: /thirdpartywarpper/managedfundshome

Disallow: /thirdpartywrapper/marketsdatahome

Disallow: /thirdpartywrapper/alertshubprefcenter

Disallow: /readitlaterwrappernorr

Disallow: /readitlaterwrapper11

Disallow: /alertshub-wrapper

Disallow: /mba-newslines-wrapper11

User-agent: *

Disallow:

The Guardian
Original URL: http://www.theguardian.com/robots.txt

guardian.txt

User-agent: *

Disallow: /sendarticle/

Disallow: /Users/

Disallow: /users/

Disallow: /*/print$

Disallow: /email/

Disallow: /contactus/

Disallow: /share/

Disallow: /websearch

Disallow: /*?commentpage=

Disallow: /whsmiths/

Disallow: /external/overture/

Disallow: /discussion/report-abuse/*

Disallow: /discussion/report-abuse-ajax/*

Disallow: /discussion/comment-permalink/*

Disallow: /discussion/report-abuse/*

Disallow: /discussion/user-report-abuse/*

Disallow: /discussion/handlers/*

Disallow: /discussion/your-profile

Disallow: /discussion/your-comments

Disallow: /discussion/edit-profile

Disallow: /discussion/search/comments

Disallow: /search

Disallow: /music/artist/*

Disallow: /music/album/*

Disallow: /books/data/*

Disallow: /settings/

Disallow: /embed/

Disallow: /*styles/js-on.css$

Disallow: /sport/olympics/2008/events/*

Disallow: /sport/olympics/2008/medals/*

User-Agent: Googlebot-Mobile

Disallow: /

Sitemap: http://www.theguardian.com/newssitemap.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-1998.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-1999.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2000.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2001.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2002.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2003.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2004.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2005.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2006.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2007.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2008.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2009.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2010.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2011.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2012.xml

Sitemap: http://spiderbytes.theguardian.com/sitemap/sitemap-2013.xml

Sitemap: http://www.theguardian.com/videositemap.xml

Sitemap: http://spiderbytes.theguardian.com/editionalised-sitemap.xml

The Independent
Original URL: http://indpendent.co.uk/robots.txt

independent.txt

##ACAP version=1.0

Site contents Copyright Independent New and Media Limited

Please note our terms and conditions http://www.independent.co.uk/service/legal-terms-amp-policies-759573.html

Permissions to make copies of site content may be obtained from the Newspaper Licensing Agency http://www.nla.co.uk

Allow all

User-agent: Mediapartners-Google

Allow: /

User-agent: *

Disallow: /*?cmp

Disallow: /*?CMP

Disallow: /*?pageToolsFontSize

Disallow: /*?printService=print

Disallow: /*ILC-refresh

Disallow: /search/

Disallow: /template/

Disallow: *printable=1

Disallow: /clickthrough.do

Disallow: /hei-fi/

Disallow: /config/

Disallow: /independentplus/

Allow all

ACAP-crawler: *

User-agent: *

ACAP-disallow-crawl: /search/

Disallow: /search/

Sitemap files

Sitemap: http://www.independent.co.uk/independent.co.uk/sitemap/sitemap.xml

Sitemap: http://www.independent.co.uk/googlenewssitemap.jsp

Sitemap: http://www.independent.co.uk/independent.co.uk/sitemap/sitemap_day-in-a-page.xml.gz

The Telegraph
Original URL: http://www.telegraph.co.uk/robots.txt

telegraph.txt

Robots.txt file

All robots will spider the domain

User-agent: *

Disallow: */ixale/

Disallow: /core/Content/

Disallow: /*?source=rss

Disallow: /*?mobile=true

Disallow: /*?mobile=basic

Disallow: /*?service=artBody

Disallow: /promotions/emails/

Disallow: /search/*

Disallow: /searchbeta/*

Disallow: /travel/8711559/The-Telegraph-Travel-Awards-2011.html

Disallow: /travel/hotel/e/*

Disallow: /sponsored/staging/

Disallow: /sponsored/email/

Disallow: /sponsored/rbth/

Disallow: /sponsored/business/national-business-awards/

Disallow: /sponsored/business/vistaprint-business-printing-se/

Disallow: /sponsored/business/youth-business-enterprise/

Disallow: /sponsored/lifestyle/kaleidoscope/

Disallow: /sponsored/motoring/alfa-romeo-cars/

Disallow: /sponsored/motoring/driving-routes/

Disallow: /sponsored/motoring/vw-up/

Disallow: /sponsored/property/damp-proofing/

Disallow: /sponsored/property/strutt-and-parker/

Disallow: /sponsored/sport/fa-community-awards/

Disallow: /sponsored/sport/juniorgolf/

Disallow: /sponsored/sport/unibet-sporting-odds/

Disallow: /sponsored/sport/walking-wounded-2013/

Disallow: /sponsored/supplement-portfolio/

Disallow: /sponsored/travel/adventuretravellive/

Disallow: /sponsored/travel/brittany_ferries/

Disallow: /sponsored/travel/caymanislands/

Disallow: /sponsored/travel/cruise-holidays/

Disallow: /sponsored/travel/cruise-travel-show/

Disallow: /sponsored/travel/dubai-family-holidays/

Disallow: /sponsored/travel/dunhill-travel-deals/

Disallow: /sponsored/travel/flanders-coast/

Disallow: /sponsored/travel/france_show/

Disallow: /sponsored/travel/great-rail-journeys/

Disallow: /sponsored/travel/guided-holiday-tours/

Disallow: /sponsored/travel/hilton-hotels-resorts/

Disallow: /sponsored/travel/holiday-movies/

Disallow: /sponsored/travel/p-and-o-cruises/

Disallow: /sponsored/travel/telegraph-cottages/

Disallow: /sponsored/travel/uk-holidays/

Disallow: /sponsored/finance/spread-betting/

Disallow: /sponsored/travel/ultimate-australia/

Disallow: /sponsored/travel/free-wheel/

Disallow: /sponsored/motoring/ford-future-sessions/

Disallow: /sponsored/entertainment/book-club-tour/

Disallow: /sponsored/travel/australia-destinations/

Disallow: /sponsored/finance/retirement-annuity/

Disallow: /sponsored/travel/hidden-britain/

Sitemap: http://www.telegraph.co.uk/newsSitemapIndex.xml

Sitemap: http://www.telegraph.co.uk/section_sitemap_index.xml

User-Agent: endeca

Disallow: /archive/

User-Agent: Mediapartners-Google

Allow: /

The Daily Mirror
Original URL: http://www.mirror.co.uk/robots.txt

themirror.txt

Sitemap: http://www.mirror.co.uk/map_news.xml

Sitemap: http://www.mirror.co.uk/sitemaps/sitemap_index.xml

User-agent: *

Crawl-delay: 10.0

Disallow: /search/

Disallow: /comm-part-test/

Disallow: /*service=ajax

Disallow: /centenary-fund/

Disallow: /3am/weird-celeb-news/xxx-1341448

Disallow: /3am/weird-celeb-news/tamara-ecclestone-watched-boyfriends-sex-1341448

Disallow: /comm-part-test/oh-like-beside-seaside-uk-1722203

User-agent: daumoa

Disallow: /

User-agent: Sosospider

Disallow: /

User-agent: rogerbot

Disallow: /

#Agent Specific Disallowed Sections

The Sun
Original URL: http://www.thesun.co.uk/robots.txt

thesun.txt

##ACAP version=1.0

#Last updated: 26/11/2007

#The Sun Online Newspaper

#Site contents copyright 2007 News Group Newspapers Ltd.

#Please note our terms and conditions - see

#http://www.thesun.co.uk/section/0,,22,00.html

#Spidering is not allowed by our terms and conditions

#Authorised spidering is subject to permission

#For authorisation please contact us - see

#http://www.nisyndication.com/about_us.html

User-agent: *

Sitemap: http://www.thesun.co.uk/newssitemapindex.xml

Sitemap: http://www.thesun.co.uk/sitemapindex.xml

Disallow: /popupWindow/

Disallow: /blockdisplay/

Disallow: /thumbnail/

Disallow: /sun.css/

Disallow: /_nipd/

Disallow: /genads/

Disallow: /_gatools/

Disallow: /cgi-bin/

Disallow: /sol/homepage/news/polish_sun/

Disallow: /sol/homepage/advertisement_feature/showbiztest/

Disallow: /search/

Disallow: /sol/homepage/news/article2196181.ece

Disallow: /sol/homepage/news/article2198458.ece

Disallow: /sol/homepage/feeds/iphone/

Disallow: /sol/homepage/feeds/iPad/

Disallow: /sol/homepage/feeds/smartphone/

Disallow: /sol/homepage/feeds/tescoFeeds/

Disallow: /sol/homepage/smartphonecolumnists/

Disallow: /*?mob=1

Disallow: /*?iPadApp=true

Disallow: /*?iPhoneApp=true

#Agent Specific Disallowed Sections

User-Agent: Googlebot-Video

Disallow: /sol/homepage/sport/football/

Disallow: /sol/homepage/showbiz/

User-agent: NewsNow

Disallow: /

User-agent: Omgili

Disallow: /

User-agent: WebVac

Disallow: /

User-agent: WebZip

Disallow: /

User-agent: psbot

Disallow: /

User-agent: ia_archiver

Disallow: /

User-agent: Meltwater

Disallow: /

#Last updated: 26/11/2007

#The Sun Online Newspaper

#Site contents copyright 2007 News Group Newspapers Ltd.

#Please note our terms and conditions - see

#http://www.thesun.co.uk/section/0,,22,00.html

#Spidering is not allowed by our terms and conditions

#Authorised spidering is subject to permission

#For authorisation please contact us - see

#http://www.nisyndication.com/about_us.html

ACAP-crawler: *

User-agent: *

ACAP-disallow-crawl: /popupWindow/

Disallow: /popupWindow/

ACAP-disallow-crawl: /blockdisplay/

Disallow: /blockdisplay/

ACAP-disallow-crawl: /thumbnail/

Disallow: /thumbnail/

ACAP-disallow-crawl: /sun.css/

Disallow: /sun.css/

ACAP-disallow-crawl: /_nipd/

Disallow: /_nipd/

ACAP-disallow-crawl: /genads/

Disallow: /genads/

ACAP-disallow-crawl: /_gatools/

Disallow: /_gatools/

ACAP-disallow-crawl: /cgi-bin/

Disallow: /cgi-bin/

ACAP-disallow-crawl: /sol/homepage/news/polish_sun/

Disallow: /sol/homepage/news/polish_sun/

ACAP-disallow-crawl: /sol/homepage/advertisement_feature/showbiztest/

Disallow: /sol/homepage/advertisement_feature/showbiztest/

ACAP-disallow-crawl: /search/

Disallow: /search/

ACAP-disallow-crawl: /sol/homepage/news/article2196181.ece

ACAP-disallow-crawl: /sol/homepage/news/article2198458.ece

The Times
Original URL: http://www.thetimes.co.uk/robots.txt

thetimes.txt

#Robots.txt File

#Version: 0.1

#Last updated: 23/05/2010

#Site contents Copyright Times Newspapers Ltd

#Please note our terms and conditions
http://www.timesonline.co.uk/tol/tools_and_services/services/terms_and_conditions/

#Spidering is not allowed by our terms and conditions

#Authorised spidering is subject to permission

#For authorisation please contact us - see http://www.nisyndication.com/contact.html

User-agent:*

Sitemap: http://www.thetimes.co.uk/newssitemapindex.xml

Sitemap: http://www.thetimes.co.uk/articlesitemapindex.xml

Sitemap: http://www.thetimes.co.uk/sectionsitemapindex.xml

Disallow: /tto/papers/

Disallow: /tto/papers.do

Disallow: /tto/feeds/

Disallow: /tto/adtest/

Disallow: /tto/public/sitesearch.do

Disallow: /tto/public/needtoknow/

#Agent Specific Disallowed Sections

User-agent: NewsNow

Disallow: /

User-agent: Omgili

Disallow: /

User-agent: WebVac

Disallow: /

User-agent: WebZip

Disallow: /

User-agent: psbot

Disallow: /

User-agent: ia_archiver

Disallow: /

User-agent: Meltwater

Disallow: /

ACAP-crawler: *

ACAP-disallow-crawl: /tto/papers/

ACAP-disallow-crawl: /tto/papers.do

ACAP-disallow-crawl: /tto/feeds/

ACAP-disallow-crawl: /tto/adtest/

ACAP-disallow-crawl: /tto/public/sitesearch.do

ACAP-disallow-crawl: /tto/public/needtoknow/

User-agent: Twitterbot

Disallow:

Miscellaneous

Greenpeace UK
Original URL: http://greenpeace.org.uk/robots.txt

greenpeaceuk.txt

$Id: robots.txt,v 1.9.2.1 2008/12/10 20:12:19 goba Exp $

#

robots.txt

#

This file is to prevent the crawling and indexing of certain parts

of your site by web crawlers and spiders run by sites like Yahoo!

and Google. By telling these "robots" where not to go on your site,

you save bandwidth and server resources.

#

This file will be ignored unless it is at the root of your host:

Used: http://example.com/robots.txt

Ignored: http://example.com/site/robots.txt

#

For more information about the robots.txt standard, see:

http://www.robotstxt.org/wc/robots.html

#

For syntax checking, see:

http://www.sxw.org.uk/computing/robots/check.html

User-agent: *

Crawl-delay: 10

Directories

Disallow: /includes/

Disallow: /misc/

Disallow: /modules/

Disallow: /profiles/

Disallow: /scripts/

Disallow: /sites/

Disallow: /themes/

Files

Disallow: /CHANGELOG.txt

Disallow: /cron.php

Disallow: /INSTALL.mysql.txt

Disallow: /INSTALL.pgsql.txt

Disallow: /install.php

Disallow: /INSTALL.txt

Disallow: /LICENSE.txt

Disallow: /MAINTAINERS.txt

Disallow: /update.php

Disallow: /UPGRADE.txt

Disallow: /xmlrpc.php

Paths (clean URLs)

Disallow: /admin/

Disallow: /comment/reply/

Disallow: /document/

Disallow: /interests/

Disallow: /filter/tips/

Disallow: /logout/

Disallow: /node/add/

Disallow: /search/

Disallow: /skills/

Disallow: /taxonomy/

Disallow: /tags/

Disallow: /user/register/

Disallow: /user/password/

Disallow: /user/login/Disallow: /image/

Paths (no clean URLs)

Disallow: /?q=admin/

Disallow: /?q=comment/reply/

Disallow: /?q=filter/tips/

Disallow: /?q=logout/

Disallow: /?q=node/add/

Disallow: /?q=search/

Disallow: /?q=user/password/

Disallow: /?q=user/register/

Disallow: /?q=user/login/

2010-12-13 JPS prevent search indexing of bare node pages

Disallow: /node/

2013-03-05 DanB - allow Twitterbot

User-agent: Twitterbot

Allow: /sites/files/gpuk/imagecache/

The Institution of Environmental Sciences
Original URL: https://www.the-ies.org/robots.txt

ies.txt

\#

\# robots.txt

\#

\# This file is to prevent the crawling and indexing of certain parts

\# of your site by web crawlers and spiders run by sites like Yahoo!

\# and Google. By telling these "robots" where not to go on your site,

\# you save bandwidth and server resources.

\#

\# This file will be ignored unless it is at the root of your host:

\# Used: http://example.com/robots.txt

\# Ignored: http://example.com/site/robots.txt

\#

\# For more information about the robots.txt standard, see:

\# http://www.robotstxt.org/wc/robots.html

\#

\# For syntax checking, see:

\# http://www.sxw.org.uk/computing/robots/check.html

User-agent: *

Crawl-delay: 10

\# Directories

Disallow: /includes/

Disallow: /misc/

Disallow: /modules/

Disallow: /profiles/

Disallow: /scripts/

Disallow: /themes/

Files

Disallow: /CHANGELOG.txt

Disallow: /cron.php

Disallow: /INSTALL.mysql.txt

Disallow: /INSTALL.pgsql.txt

Disallow: /install.php

Disallow: /INSTALL.txt

Disallow: /LICENSE.txt

Disallow: /MAINTAINERS.txt

Disallow: /update.php

Disallow: /UPGRADE.txt

Disallow: /xmlrpc.php

Paths (clean URLs)

Disallow: /admin/

Disallow: /comment/reply/

Disallow: /filter/tips/

Disallow: /logout/

Disallow: /node/add/

Disallow: /search/

Disallow: /user/register/

Disallow: /user/password/

Disallow: /user/login/

Paths (no clean URLs)

Disallow: /?q=admin/

Disallow: /?q=comment/reply/

Disallow: /?q=filter/tips/

Disallow: /?q=logout/

Disallow: /?q=node/add/

Disallow: /?q=search/

Disallow: /?q=user/password/

Disallow: /?q=user/register/

Disallow: /?q=user/login/

Gov.uk (UK government central website)
Original URL: https://www.gov.uk/robots.txt

gov.txt

User-agent: *

Disallow: /*/print

Don't allow indexing of search results

Disallow: /search?

We only allow indexing of the licence-finder landing page

Disallow: /licence-finder/

Allow: /licence-finder

We only allow indexing of the finance finder landing page

Disallow: /business-finance-support-finder/*

Allow: /business-finance-support-finder

Disallow: /apply-for-a-licence

Sitemap: https://www.gov.uk/sitemap.xml

Crawl-delay: 0.5

DEFRA
Original URL: http://www.defra.gov.uk/robots.txt

defra.txt (old)

sitemap: http://archive.defra.gov.uk/sitemap-index.xml

User-agent: *

Disallow: /cgi-bin/

Disallow: /temp

Disallow: /Templates

Environment Agency
Original URL:

EA.txt (old)

Original URL: http://www.environment-agency.gov.uk/robots.txt

User-agent: *

Disallow: /apps/flycapture/

Disallow: /common/

Disallow: /xml/

Disallow: /jobs/jobs.aspx

Disallow: /106609.aspx

Disallow: /contactus/Email_this_page.aspx

Disallow: /contactus/Feedback_on_this_page.aspx

Disallow: /cy/cysylltuani/Email_this_page.aspx

Disallow: /cy/cysylltuani/Feedback_on_this_page.aspx

Disallow: /homeandleisure/floods/132266.aspx

Disallow: /homeandleisure/floods/132278.aspx

Disallow: /homeandleisure/floods/132307.aspx

Disallow: /homeandleisure/floods/132309.aspx

Disallow: /homeandleisure/floods/132280.aspx

Disallow: /homeandleisure/floods/132292.aspx

Disallow: /homeandleisure/floods/132276.aspx

Disallow: /homeandleisure/floods/132282.aspx

Disallow: /homeandleisure/floods/132311.aspx

Disallow: /homeandleisure/floods/132313.aspx

Disallow: /homeandleisure/floods/132315.aspx

Disallow: /homeandleisure/floods/132802.aspx

Disallow: /homeandleisure/floods/132594.aspx

Disallow: /homeandleisure/floods/132284.aspx

Disallow: /homeandleisure/floods/132317.aspx

Disallow: /homeandleisure/floods/132319.aspx

Disallow: /homeandleisure/floods/fwdredirect.aspx

Disallow: /test/

Disallow: /114264.aspx

Disallow: /124623.aspx

Disallow: /124619.aspx

Disallow: /126434.aspx

Disallow: /131718.aspx

Disallow: /research/policy/132323.aspx

Disallow: /homeandleisure/recreation/boating/

Disallow: /homeandleisure/recreation/canoeing/

Disallow: /homeandleisure/133217.aspx

Disallow: /140062.aspx

Disallow: /cy/140062.aspx

Disallow: /business/140062.aspx

Disallow: /cy/busnes/140062.aspx

Disallow: /cy/test/

Disallow: /netregs/

Disallow: /cy/netregs/

Disallow: /searchtemplate.aspx

Disallow: /searchtemplate2.aspx

Disallow: /research/library/publications/40741.aspx

Disallow: /homeandleisure/floods/38317.aspx

Disallow: /business/regulation/101529.aspx

Disallow: /business/regulation/111050.aspx

Disallow: /business/regulation/31887.aspx

Disallow: /business/topics/waste/32084.aspx

Disallow: /business/topics/waste/32210.aspx

Disallow: /cy/busnes/pynciau/gwastraff/32084.aspx

Disallow: /cy/busnes/pynciau/gwastraff/32088.aspx

Disallow: /cy/busnes/pynciau/gwastraff/32210.aspx

Disallow: /cy/busnes/pynciau/gwastraff/32214.aspx

Disallow: /cy/busnes/rheoleiddio/101529.aspx

Disallow: /cy/busnes/rheoleiddio/31887.aspx

Disallow: /147983.aspx

Disallow: /shell/147983.aspx

Disallow: /147886.aspx

Disallow: /146392.aspx

Disallow: /146405.aspx

Disallow: /147804.aspx

Disallow: /147800.aspx

Disallow: /146407.aspx

Disallow: /146409.aspx

Disallow: /146411.aspx

Disallow: /146394.aspx

Disallow: /146752.aspx

Disallow: /146463.aspx

Disallow: /146396.aspx

Disallow: /146755.aspx

Disallow: /146757.aspx

Disallow: /146759.aspx

Disallow: /146761.aspx

Disallow: /146398.aspx

Disallow: /147208.aspx

Disallow: /147214.aspx

Disallow: /147217.aspx

Disallow: /147220.aspx

Disallow: /147222.aspx

Disallow: /147226.aspx

Disallow: /147224.aspx

Disallow: /147210.aspx

Disallow: /147232.aspx

Disallow: /147230.aspx

Disallow: /147228.aspx

Disallow: /147212.aspx

Disallow: /147234.aspx

Disallow: /cy/147886.aspx

Disallow: /cy/146392.aspx

Disallow: /cy/146405.aspx

Disallow: /cy/147804.aspx

Disallow: /cy/147800.aspx

Disallow: /cy/146407.aspx

Disallow: /cy/146409.aspx

Disallow: /cy/146411.aspx

Disallow: /cy/146394.aspx

Disallow: /cy/146752.aspx

Disallow: /cy/146463.aspx

Disallow: /cy/146396.aspx

Disallow: /cy/146755.aspx

Disallow: /cy/146757.aspx

Disallow: /cy/146759.aspx

Disallow: /cy/146761.aspx

Disallow: /cy/146398.aspx

Disallow: /cy/147208.aspx

Disallow: /cy/147214.aspx

Disallow: /cy/147217.aspx

Disallow: /cy/147220.aspx

Disallow: /cy/147222.aspx

Disallow: /cy/147226.aspx

Disallow: /cy/147224.aspx

Disallow: /cy/147210.aspx

Disallow: /cy/147232.aspx

Disallow: /cy/147230.aspx

Disallow: /cy/147228.aspx

Disallow: /cy/147212.aspx

Disallow: /cy/147234.aspx

Jobs.ac.uk
Original URL: http://www.jobs.ac.uk/robots.txt

jobs.ac.uk.txt

Specify Sitemap location (Sitemaps Protocol 0.9)

Sitemap: http://www.jobs.ac.uk/sitemap.xml

user-agent: *

disallow: /search

Cordis
Original URL: http:// cordis.europa.eu.robots.txt

cordis.europa.eu.robots.txt

User-agent: *

Disallow: /newsearch

Disallow: /coal-steel-rtd/survey

The Whitehouse
Original URL: http://www.whitehouse.gov/robots.txt

whitehouse.txt

#

robots.txt

#

This file is to prevent the crawling and indexing of certain parts

of your site by web crawlers and spiders run by sites like Yahoo!

and Google. By telling these "robots" where not to go on your site,

you save bandwidth and server resources.

#

This file will be ignored unless it is at the root of your host:

Used: http://example.com/robots.txt

Ignored: http://example.com/site/robots.txt

#

For more information about the robots.txt standard, see:

http://www.robotstxt.org/wc/robots.html

#

For syntax checking, see:

http://www.sxw.org.uk/computing/robots/check.html

User-agent: *

Crawl-delay: 10

Directories

Disallow: /includes/

Disallow: /misc/

Disallow: /modules/

Disallow: /profiles/

Disallow: /scripts/

Disallow: /themes/

Files

Disallow: /CHANGELOG.txt

Disallow: /cron.php

Disallow: /INSTALL.mysql.txt

Disallow: /INSTALL.pgsql.txt

Disallow: /install.php

Disallow: /INSTALL.txt

Disallow: /LICENSE.txt

Disallow: /MAINTAINERS.txt

Disallow: /update.php

Disallow: /UPGRADE.txt

Disallow: /xmlrpc.php

Paths (clean URLs)

Disallow: /admin/

Disallow: /comment/reply/

Disallow: /filter/tips/

Disallow: /logout/

Disallow: /node/add/

Disallow: /search/

Disallow: /user/register/

Disallow: /user/password/

Disallow: /user/login/

Paths (no clean URLs)

Disallow: /?q=admin/

Disallow: /?q=comment/reply/

Disallow: /?q=filter/tips/

Disallow: /?q=logout/

Disallow: /?q=node/add/

Disallow: /?q=search/

Disallow: /?q=user/password/

Disallow: /?q=user/register/

Disallow: /?q=user/login/

Appendix 3: Robots.txt results

Robots.txt findings:

The tables below show the most general characteristics of the gathered robots.txt files.

Business	Crawl Delay (seconds)	Restricted folders?	Banned agents?	Sitemap?
British Chambers of Commerce	90	No	No	No
CBI	None	Yes	No	No
BP	None	Yes	No	No
Vodafone	None	Yes	No	Yes
GlaxoSmithKline	None	Yes	No	Yes
Sabmiller	None	Yes	No	No
Rio Tinto Group	None	Yes	No	No

FIGURE 5

BUSINESS CATEGORY ROBOTS.TXT FILES

Media	Crawl Delay (seconds)	Restricted folders?	Banned agents?	Sitemap?
BBC	None	Yes	No	Yes
Daily Mail	None	Yes	No	Yes
Daily Express	None	Yes	No	No
Financial Times	None	Yes	Yes	No
Guardian	None	Yes	No	Yes
Independent	None	Yes	Yes	Yes
Telegraph	None	Yes	Yes	Yes
Daily Mirror	10	Yes	Yes	No

The Sun	None	Yes	Yes	Yes
The Times	None	Yes	Yes	Yes

Miscellaneous	Crawl Delay (seconds)	Restricted folders?	Banned agents?	Sitemap?
Greenpeace UK	10	Yes	No	No
Institute of Environmental Sciences (IES)	10	Yes	No	No
gov.uk	0.5	Yes	No	Yes
DEFRA (old)	None	Yes	No	Yes
Environment Agency (old)	None	Yes	No	No
Jobs.ac.uk	None	Yes	No	No
Cordis	None	Yes	No	No

FIGURE 7

MISCELLANEOUS CATEGORY ROBOTS.TXT FILES

Appendix 4: Python scripts

N.B. These are indicative examples of the code – variables (such as target website, directory etc.) are purposefully hard coded in the examples below to show how the code worked as it was usually launched in batches from the command line.

grablinks.py

```python
def grablinksfunc(url_to_grab_links_from):

    import urllib

    # printlinks.py

    try:

        from HTMLParser import HTMLParser

        from urllib2 import urlopen

    except ImportError:

        from html.parser import HTMLParser

        from urllib.request import urlopen

    import sys

    class PrintLinks(HTMLParser):

        def handle_starttag(self,tag,attrs):

            if tag == 'a':

                for name,value in attrs:

                    if name == 'href': print(value)

    p = PrintLinks()

    u = urllib.request.urlopen(url_to_grab_links_from)

    data = u.read()

    # charset = u.info().getparam('charset') # Python 2

    charset = u.info().get_content_charset() # Python 3

    p.feed(data.decode(charset))

    p.close()
```

```python
    return (p)


submit = 'http://www.defra.gov.uk/news/2010/05/'

a = grablinksfunc(submit)


print (a)
```

chuck_first_record.py

```python
import sys

import unicodedata

import re

import sqlite3

import urllib.parse   # for urlencode

import urllib.request

import time


#grab first record from database


def grabrecord(field, rowid):


  connection = sqlite3.connect('enviro_agency_raw.sqlite')


  cursor = connection.cursor()
  # fields: rowid, ID, Page_Title, Date, Main_Title, Body_Raw
  sql = "SELECT " + field + " FROM PR_Data_Raw WHERE rowid=" + str(rowid)


  cursor.execute(sql)
```

```python
    data = cursor.fetchall()


    connection.commit()

    connection.close()

    return (data)




def processitem(current_row):

    current_row = str(current_row)


    # data is annoyingly returned as a single tuple so have to unpack accordingly:

    # there must be a more elegant way to do this

    ID_number = grabrecord('ID', current_row)

    ID_number = ID_number[0][0]

    Date = grabrecord('Date',current_row)

    Date = Date[0][0]

    Main_Title = grabrecord('Main_Title',current_row)

    Main_Title = Main_Title[0][0]

    Body_Raw = grabrecord('Body_Raw',current_row)

    Body_Raw = Body_Raw[0][0]



    #Print statements for testing....

    #print ("ID: " + ((str)(ID_number)))

    #print ("Date: " + ((str)(Date)))

    #print ("Main title: " + ((str)(Main_Title)))
```

```python
    #print ("Body: " + ((str)(Body_Raw)))


    total = ((str)(Main_Title)) + " " + ((str)(Body_Raw))

    #total = [ID_number,Date,Main_Title,Body_Raw] #optional tuple version of compound
string


    #total = (str)(total)


    # tidy it up!


    total2 = (re.sub(r'<[^>]*?>', '', total)) #get rid of remaining markup

    total3 = (re.sub(r'&pound;', 'Â£', total2)) #exchange pound signs

    total4 = (re.sub(r' ', ' ', total3)) # exchange nbsp with spaces


    #total5 = total4.replace("\n", "") # change remaining line breaks to spaces

    #total6 = total5.encode('ascii','ignore')

    #unicodedata.normalize('NFKD', total4).encode('ascii','ignore')


    # CANT GET RID OF UNICODE OR \n YET - SEEMS TO BE AN ISSUE WITH WAY
FILE IS STORED ENCODED.


    #print ("Complete string: ")

    #print ("***************************")


    # CHUCK IT!!


    print ("Submitting to Churnalism.com.....")
```

```python
    url = 'http://churnalism.com/api/search/'

    values = {'text' : (total4)}


    data = urllib.parse.urlencode(values)


    binary_data = data.encode('ascii')

    req = urllib.request.Request(url, binary_data)

    response = urllib.request.urlopen(req)

    json = response.read() #response text from churnalism in json format


    filedirectory = "results/environmentagency/churns/"


    file_to_save = (ID_number) + ".txt"

    filename = (filedirectory + file_to_save)

    f = open(filename, 'w')


    stringjson = str(json)

    f.write(stringjson)

    f.close()


    print ("File: " + ID_number + ".txt" + " written!")



current_row = 1718


while current_row < 1968:

    processitem(current_row)
```

```python
    current_row += 1

    nextrow = str(current_row)

    print ("Next submission- row: " + nextrow)

    print ("Pausing for breath.....")

    time.sleep(1)


print ("TASK COMPLETED")

time.sleep(2)

print  ("......")

time.sleep(2)

print ("At last.")
```

parse_db_full.py

```python
import sys # temp for development and error checking (e.g for exit)

import sqlite3

import os # for walking through the relevant file directory

import re # for stripping tags


def parse_post(file_to_parse):


    filedirectory = "results/environmentagency/2011_web_pages/"

    filename = (filedirectory + file_to_parse)

    filestub = file_to_parse[0:6] # grabs 6 digit number for unique database & URL reference.


    print (filestub)

    print (filename)
```

```python
writefile = open(filename, "r").read()


#print (writefile.split("<p class=\"intro\">"))


page_title_pos = writefile.find("<h1>")
selection_end_pos = writefile.find("id=\"rightcol\"")


page_title_pos +=4 #remove h1 leading tag
selection = writefile[page_title_pos:selection_end_pos]


page_title_end = selection.find("</h1>")
page_title = selection[:page_title_end] # page title stored


date_end = selection.find("</p>", 1)
date_pos = (date_end - 11)


date_text = selection[date_pos:date_end] #date stored


selection = selection[date_end:] #strip out already found data from parsing selection


main_title_pos = selection.find("\"intro\"")


# expecting second para end from start of current selection
```

```python
main_title_end = selection.find("</p>", 2)

main_title_pos += 8 # cut out 'intro' HTML

main_title = selection[main_title_pos:main_title_end]

selection = selection[main_title_end:] #strip out already found data from parsing selection

main_text_pos = selection.find("class=\"plain_text\"")
main_text_pos += 21
main_text_end = selection.find("/div>",1)

main_text = selection[main_text_pos:main_text_end]

# get rid of commas from page title

newpagetitle = str(page_title.replace("\"", "**"))
page_title = newpagetitle

print ("Page title: " + (page_title))

print ("Date: " + (date_text))

#get rid of commas from main title too

newtitle = str(main_title.replace("\"", "**"))
main_title = newtitle
```

```python
print ("Main title: " + (main_title))

#strip tags from main_text

newvalue = (re.sub(r'<[^>]*?>', '', main_text))
newvalue2 = (re.sub(r'&pound;', 'Â£', newvalue))

# newvalue2.replace("\\'", "'")
newvalue3 = str(newvalue2.replace("\'", "**"))

main_text = newvalue3

print ("Main text: \n" + (main_text))

connection = sqlite3.connect('enviro_agency_raw.sqlite')

cursor = connection.cursor()

sqlstatement = "INSERT INTO PR_Data_Raw (ID, Page_Title, Date, Main_Title,
Body_Raw) "

sqlvalues = "VALUES ('" + filestub + "', '" + page_title + "', '" + date_text + "', '" +
main_title + "', '" + main_text + "')"

#sqlstatement = "'INSERT INTO PR_Data_Raw (ID, Page_Title, Date, Main_Title,
Body_Raw) '"
```

```python
    #sqlvalues = "'VALUES ("' + filestub + "', "' + page_title + "', "' + date_text + "', "' +
main_title + "', "' + main_text + "')"'


    sql = (sqlstatement + sqlvalues)


    cursor.execute(sql)


    connection.commit()

    connection.close()


# WALK!


path = 'results/environmentagency/2011_web_pages/'

listing = os.listdir(path)

for infile in listing:

    print ("current file is: " + infile)

    parse_post(infile)

    print ("Done! ")
```

Appendix 5: Churnalism tables

The overall churn patterns across media outlets can be seen below:

	Total	BBC	Independent	Guardian	Daily Mail	Telegraph	The Times	Others
Significant churn (>= 500)	393	174	39	33	32	36	15	64
Major churn (>= 1000)	49	15	8	5	7	4	5	5
% Significant churn	100.00	44.27	9.92	8.40	8.14	9.16	3.82	16.28
% Majority churn	100.00	30.61	16.33	10.20	14.29	8.16	10.20	10.20

FIGURE 1

ENVIRONMENT AGENCY PRESS RELEASE CHURN

Headings	Total	BBC	Independent	Guardian	Daily Mail	Telegraph	The Times	Others
Significant churn	173	28	18	34	16	17	11	49
Major churn	18	2	3	5	3	1	0	4
% Significant churn	100.00	16.18	10.40	19.65	9.25	9.83	6.36	28.32
% Major churn	100.00	11.11	16.67	27.78	16.67	5.56	0.00	22.22

FIGURE 2

DEFRA PRESS RELEASE CHURN

Headings	Mirror	Express	The Sun	Star
Significant churn	10	7	1	0
Major churn	1	0	0	0
% Significant churn	5.78	4.05	0.58	0.00
% Major churn	5.56	0.00	0.00	0.00

FIGURE 3

VISIBILITY OF DEFRA PRESS RELEASE CHURN

Bibliography

Allo, P. (2010) Putting Information First: Luciano Floridi and the Philosophy of Information, *Metaphilosophy*, 41(3), pp. 247–254.

Ashby, W. R. (1957) *An introduction to cybernetics*, London, Chapman and Hall Ltd.

Ayer, A. J. (1990) *The Problem of Knowledge*, London, Penguin Books.

Baase, S. (2007) *A Gift of Fire: International Version: Social, Legal, and Ethical Issues for Computing and the Internet*, Pearson.

Bakardjieva, M. and Smith, R. (2001) "The Internet in Everyday Life: Computer Networking from the standpoint of the domestic user," *New Media & Society*, 3(1), pp. 67–83.

Banko, M. and Brill, E. (2001) Scaling to very very large corpora for natural language disambiguation, *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, [online] Available at: http://dl.acm.org/citation.cfm?id=1073017 (Accessed May 22, 2013).

Barnes, S. (2010) *Regulating High-Frequency Trading: An Examination of U.S. Equity Market Structure in Light of the May 6, 2010 Flash Crash*, Kindle Edition, Amazon Media EU S.à r.l.

Barron, C. (2003) A Strong Distinction between Humans and Non-Humans is no Longer Required for Research Purposes: A Debate Between Bruno Latour and Steve Fuller, *History of the Human Sciences*, 16(2), pp. 77–99, [online] Available at: http://hhs.sagepub.com/cgi/doi/10.1177/0952695103016002004 (Accessed August 13, 2012).

BBC (2010) Anonymous hacktivists say Wikileaks war to continue, *BBC News*, [online] Available at: http://www.bbc.co.uk/news/technology-11935539 (Accessed May 15, 2012).

Bedau, M. A. (2000) Philosophical Content and Method of Artificial Life, In *The Digital Phoenix: How Computers are Changing Philosophy*, Revised, Oxford, Blackwell.

Bell, D. (2001) *An Introduction to cybercultures*, London, Routledge.

Bell, F. (2010) Network theories for technology-enabled learning and social change: Connectivism and actor network theory, *International Conference on Networked Learning*, pp. 526–533.

Benkler, Y. (2006) *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, London, Yale University Press.

Benkler, Y. and Shaw, A. (2010) A Tale of Two Blogospheres : Discursive Practices on the Left and Right 1, *Society*, Citeseer, 2010(June 15), pp. 1–58, [online] Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1611312.

Berry, D. M. (2011) *The Philosophy of Software: Code and Mediation in the Digital Age*, Basingstoke, Palgrave Macmillan.

Bijker, W. E. (1997) *Of bicycles, bakelites, and bulbs: toward a theory of sociotechnical change*, Cambridge, MA, MIT Press.

Boczkowski, P. J. (1999) Mutual shaping of users and technologies in a national virtual. community, *Journal of Communication*, 49(2), pp. 86–108.

Boczkowski, P. J. and Lievrouw, L. A. (2008) Bridging STS and communication studies: Scholarship on media and information technologies, In Hackett, E., Amsterdamska, O., Lynch, M., and Wajcman, J. (eds.), *New Handbook of Science, Technology and Society*, MIT Press, pp. 949–977.

Bonacich, P. (2007) Some unique properties of eigenvector centrality, *Social Networks*, 29, pp. 555–564.

Bonjour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA, Harvard University Press.

Bosanquet, B. (1920) *Implication and Linear Inference*, London, Macmillan and co., limited.

Boshmaf, Y. and Muslukhov, I. (2012) Key challenges in defending against malicious socialbots, *Proceedings of the 5th USENIX workshop on large-scale exploits and emergent threats,* LEET.

Bostrom, N. (2003) Are We Living in a Computer Simulation?, *The Philosophical Quarterly*, 53(211), pp. 243–255.

Bowker, G. C., Baker, K., Millerand, F. and Ribes, D. (2010) Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment, In Hunsinger, J., Klastrup, L., and Allen, M. (eds.), *International Handbook of Internet Research*, Springer, pp. 97–117.

Bowley, G. (2010) Lone $4.1 Billion Sale Led to 'Flash Crash' in May, *The New York Times*, 1st October, [online] Available at: http://www.nytimes.com/2010/10/02/business/02flash.html.

Boyd, D. and Crawford, K. (2011) Six provocations for big data, pp. 1–17, [online] Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431 (Accessed November 07, 2013).

Bradley, F. H. (1914) *Essays on Truth and Reality*, Oxford, The Clarendon Press.

Bradshaw, P. (2013) *Scraping for Journalists*, Kindle, Amazon Media EU S.à r.l.

Brooke, H. (2011) *The Silent State: Secrets, Surveillance and the Myth of British Democracy*, Windmill Books.

Brunnermeier, M. K. and Pedersen, L. H. (2008) Market Liquidity and Funding Liquidity, *Review of Financial Studies*, 22(6), pp. 2201–2238, [online] Available at: http://rfs.oxfordjournals.org/cgi/doi/10.1093/rfs/hhn098 (Accessed July 14, 2014).

Bryant, L., Srnicek, N. and Harman, G. (eds.) (2011) *The Speculative Turn: Continental Materialism and Realism*, Melbourne, re.press.

Bucher, T. (2012) A Technicity of Attention: How Software 'Makes Sense,' *Culture Machine*, 13, pp. 1–23.

Bynum, T. (2010) Philosophy in the information age, *Metaphilosophy*, 41(3), pp. 420–442, [online] Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9973.2010.01651.x/full (Accessed September 27, 2012).

Bynum, T. W. and Moor, J. H. (eds.) (1998) *The Digital Phoenix: How Computers are Changing Philosophy*, Oxford, Wiley-Blackwell.

Callon, M. (1986) The Sociology of an Actor-Network: The Case of the Electric Vehicle, In Callon, M., Law, J., and Rip, A. (eds.), *Mapping the Dynamics of Science and Technology*, Macmillan, pp. 19–34.

Canavan, J. (2005) *WHITE PAPER : SYMANTEC SECURITY RESPONSE The Evolution of Malicious IRC Bots*, [online] Available at: http://www.symantec.com/avcenter/reference/the.evolution.of.malicious.irc.bots.pdf (Accessed July 15th, 2014).

Carolina, R. (2009) Why BBC Click violated the Computer Misuse Act, *Computer Weekly*, [online] Available at: http://www.computerweekly.com/opinion/Opinion-Why-BBC-Click-violated-the-Computer-Misuse-Act (Accessed September 30, 2014).

Carr, J. (2011) *Inside Cyberwarfare*, Sebastopol CA, O'Reilly Media.

Cassidy, J. (2003) *Dot.con: How America Lost its Mind and its Money in the Internet Era*, New York, Harper Perennial.

Castells, M. (2000) *The Rise of the Network Society*, 2nd Ed, Oxford, Blackwell.

Cauer, E., Mathis, W. and Pauli, R. (2000) Life and Work of Wilhelm Cauer (1900 – 1945), In *Proceedings of the Fourteenth International Symposium of Mathematical Theory of Networks and Systems*, Perpignan.

Cespa, G. and Foucault, T. (2014) Illiquidity contagion and liquidity crashes, *Review of Financial Studies*, (May), pp. 1–68, [online] Available at: http://rfs.oxfordjournals.org/content/early/2014/03/10/rfs.hhu016.short (Accessed October 28, 2014).

CFTC-SEC (2010a) *Preliminary Findings Regarding the Market Events of May 6, 2010*. [online] Available: http://www.cftc.gov/ucm/groups/public/@otherif/documents/ifdocs/opa-jointreport-sec-051810.pdf

CFTC-SEC (2010b) *Findings Regarding the Market Events of May 6, 2010*. [online] Available: http://www.sec.gov/news/studies/2010/marketevents-report.pdf

Chan, E. P. (2013) *Algorithmic Trading*, Hoboken, NJ, USA, John Wiley & Sons, Inc., [online] Available at: http://doi.wiley.com/10.1002/9781118676998.

Chen, T. (2007) An Epidemiological View of Worms and Viruses, *Annual Review of Communications*, pp. 1–5.

Cohen, F. B. (1990) *A short course on computer viruses*, ASP Press, [online] Available at: http://dl.acm.org/citation.cfm?id=561894 (Accessed November 28, 2013).

Colburn, T. and Shute, G. (2010) Abstraction, Law and Freedom in Computer Science, *Metaphilosophy*, 41(3), pp. 345–364.

Collins, H. and Yearly, S. (1992) Epistemological chicken, In Pickering, A. (ed.), *Science as practice and culture*, Chicago, University of Chicago Press, pp. 301–326.

Cooper, J.M and Hutchinson, D.S. (eds.) (1997) *Plato: Complete works*, Indianapolis ,Hackett Publishing Company.

Cooper, W. (2004) Internet Culture, In Floridi, L. (ed.), *The Blackwell Guide to the Philosophy of Computing and Information*, Oxford: Blackwell.

Copeland, B. and Proudfoot, D. (1999) Alan Turing's forgotten ideas in computer science, *Scientific American*, (April), pp. 99–103.

Copeland, J. (2001) Colossus and the dawning of the computer age, In Smith, M. and Erskine, R. (eds.), *Action This Day*, Ealing: Bantam.

Copeland, J. (2004) Computation, In *The Blackwell Guide to the Philosophy of Computing and Information*, Oxford: Blackwell.

Crnkovic, G. (n.d.) Floridi's Informational Structural Realist Basis for Infocomputational Modeling of Cognizing Agents, *mrtc.mdh.se*, [online] Available at: http://www.mrtc.mdh.se/~gdc/work/JETAI-Dodig-Crnkovic.pdf (Accessed August 22, 2014).

Dahlberg, L. (2004) Internet Research Tracings: Towards Non-Reductionist Methodology, *Journal of ComputerMediated Communication*, Blackwell Publishing Ltd, 9(3), p. 0, [online] Available at: http://dx.doi.org/10.1111/j.1083-6101.2004.tb00289.x.

Davies, N. (2009) *Flat Earth News: An Award-winning Reporter Exposes Falsehood, Distortion and Propaganda in the Global Media*, London: Vintage.

Davis, K. and Patterson, D. (2012) *The Ethics of Big Data*, Sevastopol, CA: O'Reilly Media.

Delanda, M. (2011) *Philosophy and Simulation: The Emergence of Synthetic Reason*, London: Continuum Publishing Corporation.

Deleuze, G. (1990) *The logic of sense*, New York: Columbia University Press.

Deleuze, G. and Guattari, F. (1987) *A thousand plateaus*, Minneapolis: University of Minnesota Press.

Demir, H. (ed.) (2012) *Luciano Floridi's Philosophy of Technology: Critical Reflections*, New York: Springer.

Doctorow, C. (2009) Obama's whitehouse.gov nukes 2400-line Bush robots.txt file, *Boingboing.net*, [online] Available at: http://boingboing.net/2009/01/20/obamas-whitehousegov.html (Accessed August 08, 2013).

Dodge, M. and Kitchin, Rob. (2011) Code/Space: Software and Everyday Life, Cambridge, MA, MIT Press, Kindle edition.

Dodig-Crnkovic, G. and Burgin, M. (eds.) (2011) *Information and Computation: Essays on Scientific and Philosophical Understanding of Foundations of Information and Computation*, Singapore, World Scientific Publishing Co. Pte. Ltd.

Dolowitz, D., Buckler, S. and Sweeny, F. (2008) *Researching Online (Palgrave Study Skills)*, Basingstoke, Palgrave Macmillan.

Dumbill, E. (2013) Big Data and Thought Crime: An interview with Jim Adler, *Big Data*, 1(1), pp. 10–13.

Dumbill, E. (2012a) *Big Data Now: 2012 Edition*, Sevastopol, CA, O'Reilly Media, Inc.

Dumbill, E. (2012b) *Planning for Big Data*, Sevastopol, CA, O'Reilly Media Inc.

Dummett, M. (1964) Bringing About the Past, Poidevin, R. Le and Macbeath, M. (eds.), *Philosophical Review*, JSTOR, 73(3), pp. 338–359, [online] Available at: http://www.jstor.org/stable/2183661.

Ellul, J. (1964) *The Technological Society*, Vintage.

Ferrazzi, K. (2005) *Never Eat Alone: And Other Secrets to Success, One Relationship at a Time*, Doubleday Business.

Fetzer, J. H. (2000) Philosophy and Computer Science: Reflections on the Program Verification Debate, In Bynum, T. W. and Moor, J. H. (eds.), *The Digital Phoenix: How Computers are Changing Philosophy*, Revised, Oxford, Blackwell.

Fetzer, J. H. (1988) Program verification: the very idea, *Communications of the ACM*, ACM, 31(9), pp. 1048–1063, [online] Available at: http://portal.acm.org/citation.cfm?doid=48529.48530.

Fetzer, J. H. (2004) The Philosophy of AI and its Critique, In Floridi, L. (ed.), *The Blackwell Guide to the Philosophy of Computing and Information*, Oxford, Blackwell, pp. 119–134.

Fitzgerald, B. (2008) *Drupal for Education and E-Learing: Teaching and learning in the classroom using the Drupal CMS*, Birmingham, Packt Publishing.

Floridi, L. (2008a) A Defence of Informational Structual Realism, *Synthese*, 161(2), pp. 219–53.

Floridi, L. (2007) A look into the future impact of ICT on our lives, *The Information Society*, 23(1), pp. 59–64.

Floridi, L. (ed.) (2008b) Answers to Philosophy of Information: 5 Questions., In *Philosophy of Computing and Information: 5 Questions*, Automatic Press/VIP, pp. 89–95.

Floridi, L. (2004a) Information, In Floridi, L. (ed.), *The Blackwell Guide to the Philosophy of Computing and Information*, Oxford, Blackwell, pp. 40–61.

Floridi, L. (2006) Information ethics, its nature and scope, *ACM SIGCAS Computers and Society*, 36(3), pp. 21–36, [online] Available at: http://dl.acm.org/citation.cfm?id=1111649 (Accessed November 23, 2013).

Floridi, L. (2008c) Information ethics: a reappraisal, *Ethics and Information Technology*, 10(2-3), pp. 189–204, [online] Available at: http://www.springerlink.com/index/10.1007/s10676-008-9176-4 (Accessed August 14, 2012).

Floridi, L. (1999) Information ethics: on the philosophical foundation of computer ethics, *Ethics and Information Technology*, Springer, 1(1), pp. 37–56, [online] Available at: http://www.springerlink.com/index/u3888645r51377j3.pdf.

Floridi, L. (2011) Semantic Information and the Correctness Theory of Truth, *Erkenntnis*, 74(2), pp. 147–175.

Floridi, L. (2004b) *The Blackwell Guide to the Philosophy of Computing and Information*, Floridi, L. (ed.), *New York*, *Blackwell philosophy guides*, Blackwell, p. 371, [online] Available at: http://doi.wiley.com/10.1002/9780470757017.

Floridi, L. (2010) The philosophy of information: ten years later, *Metaphilosophy*, 41(3), pp. 402–419.

Floridi, L. (2009) Web 2.0 vs. the Semantic Web: a philosophical assessment, *Episteme*, 6(1), pp. 25–37.

Floridi, L. and Sanders, J. W. (2004) On the Morality of Artificial Agents, *Minds and Machines*, 14(3), pp. 349–79.

Ford, K. M. and Hayes, P. (1991) *Reasoning Agents in a Dynamic World: The Frame Problem*, Greenwich, JAI Press.

Foster, I. and Kesselman, C. (eds.) (2003) *The Grid 2: Blueprint for a New Computing Infrastructure*, 2nd ed, Morgan Kaufmann Publishers.

Friedman, T. L. (2007) *The World is Flat: The Globalized World in the Twenty-first Century*, 2nd Revise, Penguin.

Fuller, M. (2003) *Behind the Blip: Essays on the culture of software*, New York, Autonomedia.

Fuller, M. (2005) *Media Ecologies*, Cambridge, MA, MIT Press.

Fuller, M. (ed.) (2006) *Software Studies: A Lexicon*, Cambridge, MA, MIT Press.

Galitz, W. O. (2002) *The Essential Guide to User Interface Design*, 2nd ed, New York, Wiley.

Galloway, A.R. (2004), Protocol, Cambridge, MA, MIT Press.

Gauntlett, D. (2011) *Media Studies 2.0, and Other Battles around the Future of Media Research*, Kindle edition, Amazon Media EU S.à r.l.

Gauntlett, D. (2004) Web Studies: What's New, In Gauntlett, D. and Ross, H. (eds.), *Web.Studies*, 2nd Edition, London: Arnold.

Gerhart, S. L. (2013) Do Web Search Engines Suppress Controversy?, *First Monday*, 9(1), pp. 1–19.

Glanville, R. (2002) Second Order Cybernetics, In *Encyclopaedia of Life Support Systems*. Oxford, EoLSS Publishers, available at: http://greenplanet.eolss.net/MSS/default.htm

Gordon, C. (ed.) (1980) *Power/Knowledge: Selected Interviews & Other Writings: 1972–1977*, Pantheon.

Gray, L. (2011) Fish carried up a mountain on backs of llamas to escape global warming, *The Telegraph*, 12th April, [online] Available at: http://www.telegraph.co.uk/earth/earthnews/8444032/Fish-carried-up-a-mountain-on-backs-of-llamas-to-escape-global-warming.html.

Guardian (2014) WhatsApp: Facebook acquires messaging service in $19bn deal, The Guardian, [online] https://www.theguardian.com/technology/2014/feb/19/facebook-buys-whatsapp-16bn-deal (Accessed Oct 20, 2014)

Guarini, M. (2007) Computation, Coherence, and Ethical Reasoning, *Minds and Machines*, 17(1), pp. 27–46, [online] Available at: http://link.springer.com/10.1007/s11023-007-9056-4 (Accessed August 19, 2014).

Halevy, A., Norvig, P. and Pereira, F. (2009) The unreasonable effectiveness of data, *Intelligent Systems, IEEE*, 24(2), pp. 8–12, [online] Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4804817 (Accessed May 22, 2013).

Hargittai, E., Gallo, J. and Kane, M. (2008) Cross-Ideological Discussions among Conservative and Liberal Bloggers, *Public Choice*, (134), pp. 67–86.

Harman, G. (2008) DeLanda's ontology: assemblage and realism, *Continental Philosophy Review*, Springer Netherlands, 41(3), pp. 367–383, [online] Available at: http://www.springerlink.com/index/10.1007/s11007-008-9084-7.

Harman, G. (2007a) On Vicarious Causation, Mackay, R. (ed.), *Collapse*, [Online]: http://www.learnersdictionary.com/search/transducer, 2(March 2007), pp. 187–221, [online] Available at: http://faculty.virginia.edu/theorygroup/docs/harman=vicarious-causation.pdf.

Harman, G. (2007b) On Vicarious Causation, *Collapse*, II(March 2007), pp. 171–205, [online] Available at: http://collapse-journal.com/pub_collapse2.php.

Harman, G. (2009) *Prince of Networks*, Melbourne, re.press, [online] Available at: http://www.re-press.org.

Harman, G. (2011a) The importance of Bruno Latour for philosophy, *Cultural Studies Review*, pp. 31–49, [online] Available at: http://epress.lib.uts.edu.au/ojs/index.php/csrj/article/view/2153 (Accessed August 13, 2012).

Harman, G. (2011b) *The Quadruple Object*, Oxon, Zero Books.

Harman, G. (2011c) The Road to Objects, *continent.*, 1, pp. 171–179, [online] Available at: http://www.continentcontinent.cc/index.php/continent/article/viewArticle/48 (Accessed October 16, 2014).

Harraway, D. (1991) *Simians, cyborgs and women: The reinvention of nature*, New York, Routledge.

Hector, G. (1990) Where Did All Those Billions Go?, *Fortune*, 10th September.

Helland, P. (2011) If You Have Too Much Data, then "Good Enough" Is Good Enough, *Queue*, (Ddl), pp. 1–11, [online] Available at: http://dl.acm.org/citation.cfm?id=1988603 (Accessed May 22, 2013).

Hemenway, K. and Calishain, T. (2004) *Spidering Hacks: 100 Industrial-Strength Tips & Tools*, Sebastopol, O'Reilly Media Inc.

Hendricks, V. F. (2010) Knowledge transmissibility and pluralistic ignorance: a first stab vincent, *Metaphilosophy*, 41(3), pp. 279–291.

Hertz, G. and Parikka, J. (2012) Zombie Media: Circuit Bending Media Archaeology into an Art Method, *Leonardo*, 45(5), pp. 424–430, [online] Available at: http://www.mitpressjournals.org/doi/abs/10.1162/LEON_a_00438.

Holz, T. (2005) A short visit to the bot zoo, *Security & Privacy, IEEE*, pp. 76–79, [online] Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1439508 (Accessed November 27, 2013).

Hughes, T. P., Pinch, T. J. and Bijker, W. E. (1989) *The social construction of technological systems: new directions in the sociology and history of technology*, Cambridge, MA, MIT Press.

IIROC (2010) *Review of the Market Events of May 6 , 2010*, [online] Available at: http://www.iiroc.ca/Documents/2010/C6CB526F-9325-4DED-9F4F-98CDCB935FF3_en.pdf.

Incapsula, (2014) 2014 Bot Traffic Report: Just the Droids You were Looking for, *Incapsula,* [online] Available at: https://www.incapsula.com/blog/bot-traffic-report-2014.html (Accessed Jan 5, 2016).

Introna, L. D. and Nissenbaum, H. (2000) Shaping the web: why the politics of search engines matters., *The Information Society*, 16(3), pp. 169–85.

Johnson, D. G. (2004) Computer Ethics, In Floridi, L. (ed.), *The Blackwell Guide to the Philosophy of Computing and Information*, Oxford,Blackwell, pp. 65–75.

Johnson, D., Regan, P. and Wayland, K. (2010) Campaign Disclosure, Privacy and Transparency, *Wm. & Mary Bill Rts. J.*, 19(4), [online] Available at: http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/wmbrts19&section=34 (Accessed July 25, 2013).

Johnson, N., Zhao, G., Hunsader, E., Meng, J., Ravindar, A., Carran, S. and Tivnan, B. (2012) Financial Black Swans Driven by Ultrafast Machine Ecology, *SSRN Electronic Journal*, [online] Available at: http://www.ssrn.com/abstract=2003874.

Keen, A. (2012) *Digital Vertigo: How Today's Online Social Revolution Is Dividing, Diminishing, and Disorienting Us*, Constable.

Kelly, K. (2010) *What technology wants*, New York, Penguin.

Kimbell, L. (2013) The Object Strikes Back: An interview with Graham Harman, *Design and Culture*, 5(1), pp. 1–12.

King, D. (1997) *Kasparov v Deeper Blue*, London, B.T. Batsford Ltd.

Kirilenko, A., Kyle, A., Samadi, M. and Tuzun, T. (2011) The flash crash: The impact of high frequency trading on an electronic market, *Manuscript, U of Maryland*.

Kirkpatrick, D. (2011) *The Facebook Effect: The Real Inside Story of Mark Zuckerberg and the World's Fastest Growing Company: The Inside Story of the Company That is Connecting the World*, Chatham, Virgin Books.

Kolay, S., D'Alberto, P., Dasdan, A. and Bhattacharjee, A. (2008) A larger scale study of robots.txt, In *Proceedings of the 17th international conference on World Wide Web*, New York, ACM, pp. 1171–1172.

Korzybski, A. (1995) *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*, 5th ed, Institute of General Semantics.

Kushner, D. (2005) On the Internet, Nobody Knows You're a Bot, *Wired*, Issue 13.09

Kvanvig, J. (2003) *The Value of Knowledge and the Pursuit of Understanding*, Cambridge, Cambridge University Press.

Landow, G. P. (1994) *Hypertext Theory*, Baltimore, John Hopkins University Press.

Lanier, J. (2013) *Who Owns The Future?*, London, Allen Lane.

Lanier, J. (2011) *You Are Not A Gadget: A Manifesto*, London, Penguin.

Latour, B. (2007) Beware , your imagination leaves digital traces, *Times Higher Education Supplement*, 6th April.

Latour, B. (2010) Coming out as a philosopher, *Social Studies of Science*, 40(4), pp. 599–608, [online] Available at: http://sss.sagepub.com/cgi/doi/10.1177/0306312710367697 (Accessed March 20, 2012).

Latour, B. (2004) How to Talk About the Body? the Normative Dimension of Science Studies, Akrich, M. and Berg, M. (eds.), *Body Society*, TCS Centre, 10(2-3), pp. 205–229, [online] Available at: http://bod.sagepub.com/cgi/doi/10.1177/1357034X04042943.

Latour, B. (1999) *Pandoras Hope: Essays on the Reality of Science Studies*, London, Harvard University Press.

Latour, B. (2005) *Reassembling the Social*, Oxford, Oxford University Press.

Latour, B. (1987) *Science in Action: How to Follow Scientists and Engineers Through Society*, Callon, M. and Latour, B. (eds.), *Textes à lappui*, Harvard University Press.

Latour, B. (1991) Technology is Society Made Durable, In Law, J. (ed.), *A Sociology of Monsters? Essays on Power, Technology and Domination*, New York, Routledge, pp. 103–131.

Latour, B. (1993a) *The Pasteurization of France*, New Edition, Harvard University Press.

Latour, B. (1993b) *We Have Never Been Modern*, Cambridge, MA, Harvard University Press.

Latour, B. and Venn, C. (2002) Morality and Technology: The End of the Means, *Theory, Culture & Society*, 19(5-6), pp. 247–260, [online] Available at: http://tcs.sagepub.com/cgi/doi/10.1177/026327602761899246 (Accessed July 19, 2014).

Latour, B. and Woolgar, S. (1986) *Laboratory Life: The Construction of Scientific Facts*, Princeton, NJ, Princeton University Press.

Law, J. (2009) Actor Network Theory and Material Semiotics, Turner, B. S. (ed.), *The New Blackwell Companion to Social Theory*, Wiley-Blackwell, (April), pp. 141–158, [online] Available at: http://www.heterogeneities.net/publications/Law-ANTandMaterialSemiotics.pdf.

Law, J. (2004) *After Method: Mess in Social Science Research*, New York, Routledge.

Law, J. (2002) *Aircraft Stories: Decentering the Object in Technoscience.*, *Science and cultural theory*, *Science and cultural theory*, Durham, NC, Duke University Press, [online] Available at: http://eprints.lancs.ac.uk/3456/.

Law, J. and Bijker, W. E. (1992) *Shaping technology/building society: studies in sociotechnical change*, Cambridge, MA, MIT Press.

Law, J. and Hassard, J. (1999) *Actor Network Theory and After*, Law, J. and Hassard, J. (eds.), *Sociological review monograph series*, *Sociological Review Monographs*, Blackwell/Sociological Review, p. 256 p., [online] Available at: http://www.amazon.co.uk/dp/0631211942.

Lee, B., Cheng, S.-F. and Koh, A. (2010) Would Position Limits Have Made any Difference to the 'Flash Crash' on May 6, 2010, *SSRN Electronic Journal*, pp. 385–419, [online] Available at: http://www.ssrn.com/abstract=1701078.

Lessig, L. (1999), Code and other laws of cyberspace, New York, Basic Books .

Lessig, L. (2006), *Code: Version 2.0*, New York, Basic Books.

Lewis, D. (1976) The Paradoxes of Time Travel, Campbell, H. (ed.), *American Philosophical Quarterly*, JSTOR, 13(2), pp. 145–152, [online] Available at: http://www.jstor.org/stable/20009616.

Lewis, J., Williams, A. and Franklin, B. (2006) The quality and independence of British journalism, Cardiff, Cardiff University.

Lopes, D. M. (2004) Digital Art, In Floridi, L. (ed.), *The Blackwell Guide to the Philosophy of Computing and Information*, Oxford, Blackwell.

Loukides, M. (2012) *What is Data Science?*, Sevastopol, CA, O'Reilly Media Inc.

Macreary, D. and Kelly, A. (2013) *Making Sense of NoSQL*, Greenwich, CT, Manning.

Madhavan, A. (2011) Exchange-traded funds, market structure and the flash crash, *SSRN Electronic Journal*, [online] Available at: http://business.nd.edu/uploadedFiles/Academic_Centers/Study_of_Financial_Regulation/pdf _and_documents/Exchange-Traded Funds, Market Structure and the Flash Crash by Ananth Madhavan from Blackrock, Inc.pdf (Accessed October 28, 2014).

Mainzer, K. (2004) System: An introduction to Systems Science, In *The Blackwell Guide to the Philosophy of Computing and Information*, Blackwell, pp. 28–39.

Manovich, L. (2000), The Language of New Media . Cambridge, MA, MIT Press.

May, C. (2002) *The Information Society: A Sceptical View*, Cambridge, Polity Press.

May, T. (2005) *Gilles Deleuze: An Introduction*, Cambridge, Cambridge University Press.

Mayer-Schonberger, V. and Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*, John Murray.

McCallum, Q. E. (2012) *Bad Data Handbook*, Sevastopol, CA, O'Reilly Media Inc.

McTaggart, J. M. E. (1908) The Unreality of Time, Poidevin, R. Le and Macbeath, M. (eds.), *Mind*, *New Series*, Mind Association; Oxford University Press, 17(68), pp. 457–474, [online] Available at: http://www.jstor.org/stable/2248314.

Mehta, N. (2011) Credit Suisse Plans Market for Long-Term Investors, *Bloomberg*, [online] Available at: http://www.bloomberg.com/news/2011-01-11/credit-suisse-plans-light-pool-stock-market-aimed-at-long-term-investors.html (Accessed October 15, 2013).

Meillassoux, Q. (2010) *After finitude: An essay on the necessity of contingency*, London, Bloomsbury Academic.

Menkveld, A. and Yueshen, B. (2013) Anatomy of the flash crash, *Available at SSRN 2243520*, [online] Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2243520 (Accessed October 28, 2014).

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. a and Aiden, E. L. (2011) Quantitative analysis of culture using millions of digitized books., *Science (New York, N.Y.)*, 331(6014), pp. 176–82, [online] Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3279742&tool=pmcentrez&rendertype=abstract (Accessed May 21, 2013).

De Millo, R. A., Lipton, R. J. and Perlis, A. J. (1979) Social processes and proofs of theorems and programs, *Communications of the ACM*, ACM, 22(5), pp. 271–280, [online] Available at: http://portal.acm.org/citation.cfm?doid=359104.359106.

Montfort, N., Baudoin, P., Bell, J., Bogost, I., Douglass, J., Marino, M., Mateas, M., Reas, C., Sample, M. and Vawter, N. (2013) *10 PRINT CHR$(205.5+RND(1)); : GOTO 10*, Cambridge, MA, MIT Press.

Moor, J. H. (1985) What is computer ethics?, *Metaphilosophy*, 16(4), pp. 266–75.

Moore, G. E. (1939) Proof of an External World, *Proceedings of the British Academy*, (25), pp. 273–300.

Morozov, E. (2012) *The Net Delusion: How not to liberate the world*, London, Penguin.

Morozov, E. (2013) *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems that Don't Exist*, London, Allen Lane.

Mumford, L. (1967) *The Myth of the Machine: Technics and Human Development*, San Diego, Harcourt Brace and World.

Narang, R. K. (2013) *Inside the Black Box: A Simple Guide to Quantitative and High-Frequency Trading*, 2nd ed, Oxford, Wiley.

Naughton, J. (2006) Blogging and the emerging media ecosystem, *Reuters Fellowship seminar, ...*, pp. 1–10, [online] Available at: http://reutersinstitute.politics.ox.ac.uk/fileadmin/documents/discussion/blogging.pdf (Accessed August 30, 2012).

Norman, D. (1998) *The Design of Everyday Things*, Cambridge, MA, MIT Press.

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J. and Barabási, a-L. (2007) Structure and tie strengths in mobile communication networks., *Proceedings of the National Academy of Sciences of the United States of America*, 104(18), pp. 7332–6, [online] Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1863470&tool=pmcentrez&rendertype=abstract (Accessed May 22, 2013).

Oram, A. and Wilson, G. (eds.) (2007) *Beautiful Code*, Sevastopol, CA, O'Reilly Media Inc.

Paccagnella, L. (1997) Getting the Seats of Your Pants Dirty: Strategies for Ethnographic Research on Virtual Communities, *Journal of ComputerMediated Communication*, Blackwell Publishing Ltd, 3(1), p. 0, [online] Available at: http://jcmc.indiana.edu/vol3/issue1/paccagnella.html.

Paddrik, M., Hayes, R. and Todd, A. (2012) An agent based model of the E-Mini S&P 500 applied to Flash Crash analysis, *... & Economics (CIFEr ...*, [online] Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6327800 (Accessed October 30, 2014).

Park, H. W. and Thelwall, M. (2003) Hyperlink Analyses of the World Wide Web: A Review, *Journal of Computer-Mediated Communication*, 8(4).

Patil, D. J. (2012) *Data Jujitsu*, Sevastopol, CA, O'Reilly Media Inc.

Patterson, S. (2012) *Dark Pools: The rise of A.I. trading machines and the looming threat to Wall Street*, Kindle edi, Cornerstone Digital.

Penrose, R. (1999) *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford, Oxford paperbacks.

Pickering, A. (2009) the Politics of Theory, *Journal of Cultural Economy*, 2(1), pp. 197–212, [online] Available at: http://www.informaworld.com/openurl?genre=article&doi=10.1080/17530350903064204&magic=crossref.

Pollock, J. L. (1999) Procedural Epistemology – At the Interface of Philosophy and AI, In *The Blackwell Guide to Epistemology*, Malden, MA, Blackwell.

Porter, D. (ed.) (1997) *Internet Culture*, New York, Routledge.

Quine, W. (1994) Two Dogmas of Empiricism, In *From a Logical Point of View*, Cambridge, MA, Harvard University Press, pp. 20–46.

Rheingold, H. (2003) *Smart Mobs: The Next Social Revolution*, New edition, New York, Perseus Books.

Rheingold, H. (1993) *The Virtual community. Homesteading on Electronic Frontier*, Boston, Addison-Wesley.

Rhodes, B. and Goerzen, J. (2010) *Foundations of Python Network Programming*, Second, Berkeley, CA, Apress, [online] Available at: http://www.springerlink.com/index/10.1007/978-1-4302-3004-5.

Rigby, B. (2008) *Mobilizing Generation 2.0: A Practical Guide to Using Web 2.0: Technologies to Recruit, Organize and Engage Youth*, San Francisco, Jossey-Bass.

Van Rijmenam, M. (2013) Why the 3V's are not sufficient to describe big data, *Big Data Startup*, [online] Available at: http://www.bigdata-startups.com/3vs-sufficient-describe-big-data.

Rogers, R. (2006) *Information Politics on the Web*, New Ed, Cambridge, MA, MIT Press.

Rogers, R. (2000) *Preferred Placement: Knowledge Politics on the Web*, Maastricht, Jan Van Eyck Akademie.

Roush, S. (2010) The value of knowledge and the pursuit of survival, *Metaphilosophy*, 41(3), pp. 255–278.

Russell, B. (2001) The Problems of Philosophy, Oxford, OUP Oxford.

Salmon, F. and Stokes, J. (2010) Algorithms Take Control of Wall Street, *Wired*, [online] Available at: http://www.wired.com/2010/12/ff_ai_flashtrading/all/.

Saluzzi, J. (2009) *High Frequency Trading: Red Flags and Drug Addiction*, [online] Available: http://www.themistrading.com/article_files/0000/0465/061809AdvancedTrading.pdf

Savage, M. and Burrows, R. (2007) The Coming Crisis of Empirical Sociology, *Sociology*, 41(5), pp. 885–899, [online] Available at: http://soc.sagepub.com/cgi/doi/10.1177/0038038507080443 (Accessed May 22, 2013).

Schäfer, M. T. (2008) *Bastard Culture! User participation and the extension of cultural industries*, Utrecht, All Print.

Schmitz Weiss, a. and Domingo, D. (2010) Innovation processes in online newsrooms as actor-networks and communities of practice, *New Media & Society*, 12(7), pp. 1156–1171, [online] Available at: http://nms.sagepub.com/cgi/doi/10.1177/1461444809360400 (Accessed July 23, 2012).

Schrenk, M. (2012) *Webbots, Spiders, and Screen Scrapers*, San Fransisco, No Starch Press.

Scott, D. T. (2009) Bubble 2.0: Organized, online critique of "Web 2.0., *Rocky Mountain Communication Review*, 6(1), pp. 32–39.

Segaran, T. and Hammerbacher, J. (2009) *Beautiful data: the stories behind elegant data solutions*, Sevastopol, CA, O'Reilly Media.

Serres, M. (1982) *The Parasite*, Baltimore, John Hopkins University Press.

Sharkey, N. (2012) America's mindless killer robots must be stopped, *The Guardian*, 3rd December, [online] Available at: http://www.theguardian.com/commentisfree/2012/dec/03/mindless-killer-robots.

Sharkey, N. (2010) Saying 'No!'to Lethal Autonomous Targeting, *Journal of military ethics*, 9(4), pp. 299–313, [online] Available at: http://www.tandfonline.com/doi/abs/10.1080/15027570.2010.537903 (Accessed August 01, 2012).

Shirky, C. (2008) *Here Comes Everybody: The Power of Organizing Without Organizations*, London, Penguin Books, p. 344.

Shoemaker, S. (1969) Time Without Change, Poidevin, R. Le and Macbeath, M. (eds.), *Journal of Philosophy*, Journal of Philosophy, Inc., 66(12), pp. 363–381, [online] Available at: http://www.jstor.org/stable/2023892.

Singer, P. W. (2011) *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, Reprint ed, London, Penguin.

Sismondo, S. (2009) *An Introduction to Science and Technology Studies*, Oxford, Wiley-Blackwell.

Smith, A. (2003) Do you believe in ethics? Latour and Ihde in the trenches of the science wars (or: Watch out, Latour, Ihde's got a gun), In Ihde, D. and Selinger, E. (eds.), *Chasing technoscience: Matrix for materiality*, Bloomington, Indiana University Press, pp. 182–194.

Solove, D. (2008) *The Future of Reputation: Gossip, Rumor, and Privacy on the Internet*, New Haven, Yale University Press.

Soros, G. (2008) The Crisis and What to Do About It, *The New York Review of Books*, 55.19 (2008): 4

Stahl, B. C. (2008) Discourses on information ethics: The claim to universality, *Ethics and Information Technology*, 10(2-3), pp. 97–108, [online] Available at: http://www.springerlink.com/index/10.1007/s10676-008-9171-9 (Accessed July 26, 2012).

Stanton, J. (2012) *An Introduction to Data Science*, Syracuse, Syracuse University.

Star, S. L. and Bowker, G. C. (2002) "How to Infrastructure," In Lievrouw, L. A. and Livingstone, S. (eds.), *Handbook of New Media. Social Shaping and Consequences of ICTs*, Sage, pp. 151–162.

Steele, J. and Iliinksy, N. (eds.) (2010) *Beautiful Visualisation*, Sevastopol, CA, O'Reilly Media Inc.

Steiner, C. (2013) *Automate This: How Algorithms Came to Rule Our World*, London, Portfolio Penguin.

Stirner, M. (1995) *The Ego and Its Own*, Cambridge University Press.

Stroud, B. (1984) *The Significance of Philosophical Scepticism*, Oxford, OUP Oxford.

Sun, Y., Zhuang, Z. and Giles, C. L. (2007) A large-scale study of robots.txt, In *Proceedings of the 16th international conference on World Wide Web*, New York, ACM, pp. 1123–1124.

Swinburne, R. (2000) *Epistemic Justification*, Oxford, Oxford University Press.

Swinburne, R. (1999) *Providence and the Problem of Evil*, Oxford, Oxford University Press.

Taleb, N. (2007) *The Black Swan: the impact of the highly improbable*, New York, Random House.

Tapscott, D. and Williams, A. (2008) *Wikinomics*, London, Atlantic Books.

Tenner, E. (1997) *Why things bite back: Technology and the revenge of unintended consequences*, New York, Vintage.

Thagard, P. (2000) *Coherence in Thought and Action*, New Ed edi, Cambridge, MA, MIT Press.

Tsvetovat, M. and Kouznetsov, A. (2011) *Social Network Analysis for Startups*, Sevastopol, CA, O'Reilly Media, Inc.

Turing, A. M. (1936) On computable numbers, with an application to the Entscheidungs problem, *Proceeedings of the London Mathematical Society*, 2(42), pp. 230–65.

Turkle, S. (1995) *Life on the Screen: Identity in the Age of the Internet*, New York, Simon and Schuster.

Ure, G. (2010) Decoding the Past: Doomsday 2012, *peoplenomics.com*, [online] Available at: http://peoplenomics.com/history.htm (Accessed August 01, 2014).

Urquhart, A. (2004) Complexity, In *The Blackwell Guide to the Philosophy of Computing and Information*, Oxford, Blackwell.

U.S.-Canada Power System Outage Task Force, (2004), Final Report on the August 14th Blackout in the United States and Canada, http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/BlackoutFinal-Web.pdf (Accessed Feb 24th, 2016).

Venturini, T. (2010a) Building on faults: How to represent controversies with digital methods, *Public Understanding of Science*, 21(7), pp. 796–812.

Venturini, T. (2010b) Building on faults: How to represent controversies with digital methods, *Public Understanding of Science*, 21(7), pp. 796–812.

Venturini, T. (2009) Diving in magma: how to explore controversies with actor-network theory, *Public Understanding of Science*, 19(3), pp. 258–273, [online] Available at: http://pus.sagepub.com/cgi/doi/10.1177/0963662509102694 (Accessed May 25, 2013).

Venturini, T. and Latour, B. (2010) The Social Fabric: Digital Traces and Quali-quantitative Methods, *Proceedings of Future En Seine*, [online] Available at: http://www.tommasoventurini.it/web/uploads/tommaso_venturini/TheSocialFabric.pdf (Accessed July 10, 2013).

Waddington, S. (ed.) (2012) *Share This*, Hoboken, John Wiley and Sons Ltd.

Wallach, W. and Allen, C. (2010) *Moral Machines*, New York, OUP USA.

Wang, W. (2006) *Steal This Computer Book 4.0: What They Won't Tell You About the Internet*, San Francisco, No Starch Press.

Watts, D. J. (1999) *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton University Press.

Welchman, A. (1997) Machinic thinking, In Ansell-Pearson, K. (ed.), *Deleuze and Philosophy*, Routledge, pp. 211–229, [online] Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Machinic+Thinking#0 (Accessed June 01, 2012).

Wellman, M. P., Greenwald, A. and Stone, P. (2007) *Autonomous Bidding Agents*, Cambridge, MA, MIT Press.

Westin, A. F. (1968) *Privacy and Freedom*, New York, Atheneum.

Weston, D. (2010a) Could machines be responsible for a major disaster?, *European Journal of Mechanical and Environmental Engineering*, 1, pp. 32–35.

Weston, D. (2010b) Unleashing Web 2.0 for Autonomous Learners, In Corker, C. (ed.), *CPLA 2010 Proceedings*, Sheffield, Sheffield Hallam University, pp. 111–118.

Whitehead, A. N. (1947) *Essays in Science and Philosophy*, *Philosophy and Phenomenological Research*, Philosophical Library, [online] Available at: http://www.jstor.org/stable/2103405?origin=crossref.

Whitehead, A. N. (1960) *Process and Reality*, New York, Harper & Row.

Wiener, N. (1954) *The Human Use of Human Beings*, Boston, Houghton Mifflin.

Williams, B. (1973) *Problems of the Self*, Cambridge, Cambridge University Press.

Williams, M. (1999) Skepticism, In Greco, J. and Sosa, E. (eds.), *The Blackwell Guide to Epistemology*, Malden, MA, Blackwell Publishers Ltd, pp. 35–70.

Williams, M. (1996) *Unnatural Doubts: Epistemological Realism and the Basis of Skepticism*, Princeton, Princeton University Press.

Williamson, M. M. and Léveillé, J. (2003) An epidemiological model of virus spread and cleanup, *Proceedings of the Virus Bulletin Conference*.

Winner, L. (1980) Do artifacts have politics, *Daedalus*, 109(1), pp. 121–136.

Winner, L. (1988) The whale and the reactor: A personal memoir, *Journal of American Culture*, 3(448), pp. 433–449.

Winner, L. (1993) Upon opening the black box and finding it empty: Social constructivism and the philosophy of technology, *Science, Technology, & Human Values*, 18(3), pp. 362–378, [online] Available at: http://www.jstor.org/stable/10.2307/689726 (Accessed November 26, 2013).

Woolgar, S. (1991) The Turn to Technology in Social Studies of Science, *Science, Technology, & Human Values*, 16, pp. 20–50.

Woolgar, S. (ed.) (2002) *Virtual society? Technology, cyberbole, reality*, Oxford, Oxford University Press.

Zalabardo, J. L. (2006) Bonjour, Externalism and the Regress Problem, *Synthese*, 148(1), pp. 135–169.

Zittrain, J. (2009) *The Future of the Internet: And How to Stop It*, London, Penguin Books.