Threats to Autonomy from Emerging ICT’s

BRANDT DAINOW

A thesis submitted for the degree of

Doctor of Philosophy

Department of Computer Science
Maynooth University

September 2018

Supervisors:

Dr. J. Keating, Department of Computer Science, Maynooth University

Prof. B. Stahl, Centre for Computing and Social Responsibility, De Montfort University

Head of Department: Dr J. Timoney

CONTENTS

Abstract

This thesis investigates possible future threats to human autonomy created by currently emerging ICT's. Prepared for evaluation as PhD by Publication, it consists of four journal papers and one book chapter, together with explanatory material.

The ICT's under examination are drawn from the results of the ETICA project, which sought to identify emerging ICT's of ethical import. We first evaluate this research and identify elements in need of enhancement – the social aspects pertaining to ethical impact and the need to introduce elements of General Systems Theory in order to account for ICT's as socio-technical systems. The first two publications for evaluation present arguments from marxist and capitalist perspectives which provide an account of the social dimensions through which an ICT can reduce human autonomy.

There are many competing accounts of what constitutes human autonomy. These may be grouped into classes by their primary characteristics. The third publication for evaluation cross-references these classes with the ICT's identified by the ETICA project, showing which version of autonomy could be restricted by each ICT and how. Finally, this paper induces from this analysis some general characteristics which any ICT must exhibit if it is to restrict autonomy of any form.

Since ICT's all operate in the same environment, the ultimate effect on the individual is the aggregated effect of all those ICT's with which they interact and can be treated as an open system. Our fourth paper for evaluation therefore develops a theory of ICT's as systems of a socio-technical nature, titled "Integrated Domain Theory". Our fifth publication uses Integrated Domain Theory to explore the manner in which socio-technical systems can restrict human autonomy, no matter how conceived. This thesis thus offers two complementary answers to the primary research question.

Acknowledgements

I would like to thank my supervisors, Dr Keating and Professor Stahl, for their patient and continuous support, encouragement, and invaluable guidance, without which this project could not have succeeded.

I would like to thank Maynooth University for the welcome and support offered to mature students such as myself. Not all universities are so supportive of mature students, and Maynooth University's policies in this regard are noteworthy. I am also appreciative of the efforts many administrative staff have made on my behalf over the years. Their perpetually helpful and positive attitude is an important, though often neglected, factor in the success of many at this university, including myself. I also wish to thank the university for the efforts made on my behalf to develop a relationship with De Montfort University in order to facilitate the invaluable supervision by Prof. Stahl.

I am grateful to the Hume Scholarship fund, without which I could not have undertaken this research.

I wish to offer a special thanks to my High School History teacher, Mr Terry Aitken, now deceased. In addition to teaching history, he expended great efforts to teach his students a formal method of note-taking which has been invaluable throughout my life. This methodology has been fundamental to my ability to absorb and organise new information quickly. It has been critical to my success in this project and I will forever remain grateful for his efforts in forcing this methodology into my unwilling brain until I got it right.

Finally, I would like to thank my wife, Majella, who has patiently survived the emotional upheaval associated with living with a PhD student. It is doubtful that I could have completed this project without her continual encouragement.

1. Introduction - Project Summary

This section summarises the entire project, outlines the most important points relating to the choice of research question and summarises findings and the contributions to knowledge.

1.1. Preamble

As a PhD by Publication, this thesis cannot follow the same format as is taken in a traditional Research PhD. The lack of canonical guidelines for evaluation of PhD's by Publication is notable (Frick 2016). Three approaches are possible. Where publications are spread over many years, the aim is to recognise the individual's contribution to their field and so it is accepted the thesis should focus on demonstrating the impact those publications have had (such as by citation indexing) (Durling 2013). There are two possible approaches where PhD by Publication examination is limited to publications produced while registered as a PhD candidate (Badley 2009), as occurs at Maynooth. Maynooth University's guidelines are clear on format, but neutral on the assessment methodology. One approach is to produce a thesis which provides the same material as would appear in a traditional Research PhD, but cannot appear in publications, such as formal literature reviews. The aim is to produce something identical to a traditional thesis, in which the publications are used as notes on which to base the thesis. The aim in this approach is to enable assessment of the thesis in the same way, and on the same grounds, as a traditional Research PhD (Durling 2013). The alternative approach, which we[1] have taken here, is to provide material which directly supports assessment of the publications by means of contextualisation and linkage into a coherent narrative (Badley 2009). In our view, the primary focus for examiners in PhD by Publication is the evaluation of the publications, so the function of the thesis is to provide material aiding understanding and assessment of the published works. It is therefore most appropriate to provide a descriptive chronological account of the research process contextualising the publication of each paper. This means the chapters in such a thesis may not focus as tightly on the central argument as much as would be found in a traditional Research PhD. Instead, this account will provide a chronologically

[1] There is much debate regarding whether it is better to use the first person ('I') or third person ('we') in formal documents such as this. In our view, neither is better. Some literary cultures (defined by nation or publication) preference one, but there is no authoritive reasoning in preference to either. Neither is there universal agreement on the importance of this issue. Our approach has always been to follow the conventions of the publication and so our useage varied according to the style of the journal. Since three of the five publications used the third person, we have adopted the third person throughout the thesis in order to minimise the variance from published format.

organised account of the problems which arose during the research process, the steps taken to address them and the manner in which this was represented within the published works. Due to the necessarily short length of a journal publication, much of the background understanding used to produce a paper is often contained in little more than a quick reference to something already familiar to the readers. It is therefore appropriate that this thesis sometimes provide additional detail regarding the author's understanding and use of such concepts, how they are integrated into the research process, what methodological or other issues they addressed, and how they contributed to answering the research questions. However, this is appropriate if, and only if, such issues are relevant to assessment of the publication and are not apparent in the published material.

1.2. Research Questions

The primary research question is:

- "What are the possible threats to human autonomy generated by emerging ICT's?"

Answered in the following publications for consideration:

Threats to Autonomy from Emerging ICTs (Dainow 2017b) (see *Chapter 6. Autonomy under Emerging ICT's, p.112*).

Binding the Smart City Human-Digital System with Communicative Processes (Dainow in-press) (see *Chapter 8. Autonomy within the Integrated Node, p. 148*).

Secondary questions which must be answered first are:

1. What are the emerging ICT's?

Answered in the following publication for consideration:

Threats to Autonomy from Emerging ICTs (Dainow 2017b) (see *Chapter 6. Autonomy under Emerging ICT's, p.112*).

2. What is human autonomy?

Answered in the following publication for consideration:

Threats to Autonomy from Emerging ICTs (Dainow 2017b) (see *Chapter 6. Autonomy under Emerging ICT's, p.112*).

3. By what mechanism can an aspect of an emerging ICT have a restrictive effect on human autonomy?

Answered in the following publications for consideration:

Key Dialectics in Cloud Services (Dainow 2015b) (see *Chapter 4. The Current State of Affairs – Poles of the Debate, p.73*)

Digital Alienation as the Foundation of Online Privacy Concerns (Dainow 2015a) (see *Chapter 5. Digital Alienation - ICT's as Socio-Technical Systems, p. 91*)

1.3. Key terms in the research questions

1.3.1. "Emerging ICT's"

ICT's were defined as "emerging" if they were not merely theoretical, but were in prototype or early production, had not yet developed into their final form and/or usage in society, and were likely to have a significant impact on individuals or the wider society (Stahl 2013). This definition is drawn from the ETICA project, which sought to identify these emerging ICT's. From an initial examination of 107 technologies, ETICA produced a taxonomy of eleven emerging IC technologies. The ETICA project's definitions and taxonomy constitute the primary dataset upon which this research is based. The findings of the ETICA project were accepted as answering the secondary research question of "what are the emerging ICT's?"

ETICA's work is covered in detail and critically evaluated in *2.4 Dataset: The ETICA Project* (p.30) and *3. Evaluation of ETICA's Technology Analysis* (p.33).

1.3.2. Why autonomy?

This section explains why autonomy was selected as the topic of research, as opposed to any other ethical value.

It is beyond the scope of a single PhD research project to consider all the ethical issues of all emerging ICT's. It is therefore necessary to restrict the scope of research to that which will yield the most results within allowed timeframes. The justification for focusing on autonomy is its central and foundational position in modern applied ethics; underpinning the capacity to give consent (Agich 2003) and being an enabling condition for human rights, such as freedom (Cannataci 2016). Western society holds an assumption that personal autonomy is an absolute good (Chirkov, Ryan, and Sheldon 2011; Dworkin 1988). The capacity for autonomy is what gives humans dignity and the right to be treated as an end in themselves, not a means to an end (Kant 1998; Schneewind 2007; Buss and Zalta 2015). This position underpins European and related legal and ethical codes, as well as international regulatory regimes developed under international law since the Second World War (Dainow 2013). For example, the Universal Declaration of Human Rights is based on the premise that all humans have the capacity for autonomy (Marshall 2008). Consequently, autonomy provides the basis for all human rights legislation, much international law and provides the justification for democratic politics. In many countries the human capacity for autonomy is the basis for equality before the law (Dworkin 1988; Marshall 2008; Chirkov, Ryan, and Sheldon 2011; Schneewind 2007; Buss and Zalta 2015). The capacity for autonomy is thus the single point of ethical interest with the widest range of implications. Because autonomy is necessary (though not sufficient) for so many aspects of human flourishing, restrictions on human autonomy have the capacity to be serious and harmful, both to individuals and society as a whole. Consequently, the exposure of risks to autonomy before they actualise is an important and practical application of ICT ethics. However, there are competing definitions of autonomy in modern philosophy. Analysing the impact of a technology on autonomy therefore requires reference to the definition of autonomy being used. This issue is considered in depth in the third publication for consideration, *Threats to Autonomy from Emerging ICT's* (Dainow 2017b), which can be found in *Chapter 6. Autonomy under Emerging ICT's* (p.112).

Technologies alter the way humans interact with the world (Stahl 2011b; Misa 1994; Jonas 1984; Spiekermann 2015; Brey 2012) and so emerging ICT's change the conditions and possibilities for human action – the way people live. The concern motivating this research is that emerging ICT's could force people to live in ways which they would not have chosen. This is not affected by whether these ICT's provide other benefits. That they may do so is irrelevant to the research questions. We are concerned purely with identifying the characteristics of a potential threat. We are especially concerned with empirically verifiable characteristics extant today and with demonstrable trends in current innovation which have the potential to bring the threat into reality unless addressed. These trends constitute evidence that the threat is a genuine possibility and that this research is not merely a rhetorical exercise, but rather there is a demonstrable basis for the concern. This analysis of current characteristics of concern is found in the first two publications - *Key Dialectics in Cloud Services* and *Digital Alienation as the Foundation of Online Privacy Concerns*, which can be found in *Chapter 4. The Current State* *of Affairs – Poles of the Debate* (p. 73) and *Chapter 5. Digital Alienation - ICT's as Socio-Technical Systems* (p. 91) respectively.

1.3.3. Defining 'privacy' - *pros hen* ambiguity

This section outlines our use of the term 'privacy' in the publications. No publication offered a formal definition, nor was the term used in the same manner throughout all publications. The purpose of this section is to defend this treatment and to summarise how the term was used in the different publications.

It is generally accepted that the term 'privacy' is difficult to define because it has multiple legitimate interpretations. The term has different meanings in different cultures (Capurro 2009; Whitman 2004; Ess 2013), while the meaning also varies under different usages (Tavani 2007). However, 'privacy' is not unique in this respect. There are many terms which can have more than one referent. Such terms are said to have a *pros hen* ambiguity (Aristotle 2014a). The example offered by Aristotle is 'health.' Some things can be healthy, which means they promote health. Healthy things can be foods or behaviours. Health can be a non-material quality. We can have a "healthy" conversation, meaning it is experienced as robust, vital and other qualities which we also experience when our bodies are in a state of health. In such cases, exploration of the term can occur via discussions of usage or through attempts to characterise the referent. However, it is impossible to discuss the details of the referent of a *pros hen* term unless one first restricts the range of relevant referents. Some disagreements over definition occur because of disagreements regarding what the appropriate referent should be. Others occur because the competing definitions use different referents, but the protagonists have not noticed. Disagreements can also occur regarding whether one should be using analysis through usage or working from the referent.

This is the case with definitions of 'privacy.' Descriptive accounts of 'privacy' define it in terms of what it is made of. Here it is defined as information with particular attributes, most importantly that it is about an individual. Interest-based accounts define 'privacy' in terms of its relationship to people, such that people have an interest in the use of this information for some reason. Rights-based accounts follow a similar track, defining 'privacy' in terms of actions and their associated norms. This includes accounts which define privacy in terms of seclusion (Camp and Osorio 2003). These accounts are compatible with each other because they discuss different referents. They are all plausible in that they accurately reflect the manner in which the word 'privacy' is used.

The meaning of the term 'privacy' is thus contextual and so the meaning of the term varied between the different papers as appropriate. *Digital Alienation* used the term in reference to information, rights and interests. Here, alienation involved (among others things) issues of rights and interests with reference to the production and use of private information, as did the discussion of privacy protection in *Key Dialectics*. The latter sections of this paper focused on privacy in terms of information type within the context of access control, which includes, but is not limited to, issues of rights and interests. In both papers, information said to be "personal" if it pertained to a single individual. This is not a universally accepted definition. In some Asian cultures privacy is held to be a property of a group, such as a family, not the individual (Ess 2013). However, we have consistently used a Western understanding of privacy (Capurro 2009) as pertaining to the individual and being an enabling condition for dignity and liberty (Whitman 2004). This relation of privacy to the individual is consistent throughout all published works and holds that being personal information is necessary, but not sufficient, for information to be private. We have noted our adherence to Western concepts when discussing the limitations of the project.

Threats to Autonomy did not devote much attention to privacy. It was most prevalent in discussions of affective computing, with regard to the potential of affective systems to detect emotions which someone would choose to hide. This treatment is consistent with a number of accounts of privacy, including seclusion theories (the right to be secluded from others) (Hadley 1895) and access theories (control of access to private information) (Tavani 2007).

The last two publications describe Integrated Domain Theory. This theory does not treat of the individual *per se*, but works at the higher ontological level of the *integrated personage*. Most theories of privacy are compatible with, but insufficient for, privacy within the integrated domain. Here 'privacy' is maintained as a *pros hen* term and so relates to a number of elements. The "zone" of privacy is those areas of an integrated personage containing processes and data relating to the individual. Here we understand that privacy may relate to the processing of personal information as well as that information itself. We also recognise it is possible that those processes themselves may be private even if they do not process private information. For example, that someone's integrated personage contains processes for controlling blood pressure is private, though it contains no data about their blood pressure, because the existence of the process itself reveals private information. Similarly, mere knowledge of the existence of certain integrated devices may also be private, such as knowing someone has a pace maker. Thus the content, processes and composition of the integrated personage all constitute aspects of the integrated personage's privacy. Information, such as passwords, which grants access to private information may also be regarded as private if it is unique to that individual.

In all papers, a privacy "violation" is held to occur whenever private information is gathered, processed or used against or without the person's consent. Our concern is with autonomy, so a privacy violation is relevant to our analysis if, and only if, it results in a reduction of autonomy. Under many accounts of autonomy, such as those of Frankfurt, Dworkin and Meyers, failure to obtain permission does not restrict autonomy if the individual would have consented had they been given the opportunity (Frankfurt 1971; Dworkin 1988; Meyers 1989a). On this basis, it is not possible to assert that any given usage of private information is necessarily a restriction on autonomy. If the individual concerned accepts such a definition of autonomy for their own life, and would have consented had they had the chance, then their autonomy has not been violated. However, if they do not live by such a version of autonomy or would not have consented, then their autonomy has been violated.

1.3.4. Defining 'surveillance.'

Surveillance involves gathering information of any type about an individual for a purpose over time (Galič, Timan, and Koops 2017). Mere observation without an aim is not sufficient to qualify as surveillance. Surveillance contains acts of observation. Observation is an individual act of gathering data. It may, but does not necessarily, may be a portion of a process of observation. However, this depends on the usage made of the data gathered, not the nature of the act of observation itself. Some of the sources we used, such as Fuchs (Fuchs 2013) and Andrejevic (Andrejevic 2011b) hold that surveillance necessarily includes usage of this information (such as control of the subject) for observation to constitute surveillance. We did not include reference to specific forms of usage in the concept of surveillance for two reasons. Firstly, accounts which include control regard surveillance as a two-stage process of first gathering information and then using it. On this basis, theories of surveillance can be divided into those which require only goal-oriented observation, without reference to usage, and those which require some form of usage in order for it to qualify as suveillance. Both versions thus include the requirement for goal-orientation in observation. Goal-orientation is therefore agreed upon as a necessary characteristic of surveillance under all accounts. Thus, in practical terms, the treatment of the process from observation to useage is the same under both types of accounts – first observation must be described, then useage. For our purposes of analysis, it therefore makes little difference whether useage is held under the heading of 'surveillance' or something else. We can therefore accommodate the observational act shared by both accounts. When we then consider useage, it makes no difference to our analysis whether this is labelled 'surveillance' or not. Secondly, this is not to say we do not consider the nature of the goal motivating the surveillance. Since our concern is only with threats to human autonomy, we are only concerned with surveillance which will be used in a manner which has the potential to reduce human autonomy. Our concern is therefore only with surveillance where the goal is to achieve effects which could reduce human autonomy. In practical terms, the vector of most interest is personalisation services, in which device output is skewed in favour of the system's understanding of the individual. The axiom of digital scepticism dictates that the system's understanding of the individual is never complete or objective, nor can it be assumed to be accurate. In political terms, we can therefore state that surveillance is the process of gathering data about an individual through multiple acts of observation for the purposes of maintaining that individual's personal profile.

The term was used consistently throughout the first two publications, in which it referred to the tracking and recording of user activity on the internet and focused on the industrial-scale creation of personal profiles. This was included in the term in *Threats to Autonomy from Emerging ICT's* (*see p. 112*). This paper also allowed for the emergence of new possibilities for surveillance through emergent technologies, such as ambient intelligence and affective computing. The last two publications assumed surveillance would be an ambient characteristic of digital environments. Some mention was made of the necessary scale and complexity of such surveillance, and to some early developments, such as Cobham Plc's "Smart City Surveillance Architecture" (Cobham Plc 2012).

1.4. Summary of Research Process

The research can be divided into two phases. Phase One followed the original plan, which was to examine emerging ICT's individually and consider how each could threaten human autonomy. However, in answering this question two things were discovered which required a change of approach. Firstly, there are multiple contested versions of what constitutes autonomy. It was found that the concept becomes prone to several systemic problems once it is expanded from Kant's original strict definition. Modern philosophy's attempts to resolve these issues have led to competing definitions of autonomy. Secondly, empirical research into the effects of contemporary digital systems has revealed that the effects on a person do not derive from only that one technology. Because the person is exposed to multiple technologies at the same time, a more effective way of examining the problem is to treat of a *digital environment* composed of multiple technologies. Phase Two therefore explored the manner in which this digital environment may restrict autonomy. The five publications for consideration herein can therefore be grouped into these two phases. The first three publications constitute the steps towards the first answer - how do specific individual technologies impact different versions of autonomy? The last two papers seek to supersede these divisions of technologies and autonomies by developing a model of a digital environment, then suggesting ways in which that environment could restrict human autonomy, no matter how autonomy has been defined.

1.4.1. Phase One - how do specific individual technologies impact specific versions of autonomy?

It was beyond the scope of this research to independently determine what constitute today's significant emerging ICT's. Such a task is a considerable research project in its own right. This work had already been undertaken by the ETICA project, the largest and most comprehensive such analysis in technology studies, ICT ethics and related fields to date (Brey 2012). Furthermore, the findings of the ETICA project are now foundational to many more recent projects, indicating its acceptance and utility by the research community (see *3.3. ETICA's Wider Contribution, p.34*). The initial task was therefore to assess the quality of the relevant ETICA findings. The ETICA project created a taxonomy of eleven ICT technology groups and identified their key ethical issues (Stahl 2013). Our research therefore commenced with evaluation of ETICA's findings (see *3. Evaluation of ETICA's Technology Analysis, p.33*). Accepting ETICA's definition of technologies as socio-technical systems (Stahl 2011b), ETICA's analysis was found to be acceptable but incomplete, needing a more substantive account of the social forces underpinning ICT innovation and its impact in order to be useable for our purposes. Research was therefore needed in order to deepen the account of the social aspect of technologies. This was not aimed at a general description of the position of technology within the social, but focused on the mechanisms by which social forces interact with the individual in the context of ICT design and use. Even this is a large and complex area, and no attempt was made to develop a comprehensive account. Instead the aim was limited to explanatory frameworks sufficient to account for ICT processes which could impact human autonomy. This resulted in the first two publications, *Digital Alienation as the Foundation of Online Privacy Concerns* (Dainow 2015a) and *Key Dialectics in Cloud Services* (Dainow 2015b). These sought to demonstrate that ICT's could have negative ethical impacts on humans, to provide a language by which to describe this and to provide a preliminary account of the mechanisms by which ICT's could reduce human autonomy. While each paper addressed different specific issues, taken together they constitute a comprehensive answer to the secondary research question "by what mechanisms can ICT's restrict human autonomy?" The reason for using two papers to answer the one question is our desire to avoid committing ourselves to either a marxist or capitalist position. Instead, we sought to analyse the question from both perspectives. *Digital Alienation as the Foundation of Online Privacy Concerns* (Dainow 2015a) developed a marxist analysis, while *Key Dialectics in Cloud Services* (Dainow 2015b) considered the matter from a capitalist perspective. It is notable that certain features emerge as problematic under both analyses. For example, user lock-in through lack of interoperability by service providers violates the need for a free market under capitalist analysis while also constituting coercion under a marxist analysis.

In brief, the account we developed may be summarised as follows:

Our account commenced with acceptance of the following premises. The impact of technology on the person is mediated by social and biological factors within the individual. The concept of affordances (Hutchby 2003) provides the theoretical explanation of the way physical, psychological and biological elements are linked at the moment of perception (Zhang and Patel 2006), while Signal Detection Theory provides empirical evidence in support of affordances (Balakrishnan and MacDonald 2001; MacMillian 2002; Verghese 2001). Operant conditioning provides an account of the mechanisms by which user behaviour changes over time and by which affordances are formed, reinforced or extinguished (Killeen and Jacobs 2016; Nawyn, Intille, and Larson 2006). Advertising and other manipulations of symbolic capital provide the cognitive drivers controlling the attributions of value which drive the cognitive elements of affordances (Bernays 1928; Eshraghi and Harwood 2013; Hartson 2003). It is outside the scope of our research to provide arguments in support of these premises. However, they would not be considered controversial within many related fields, especially user interface and other forms of technology design, and have received empirical support

(Nawyn, Intille, and Larson 2006). This account is most formally represented in the second publication for consideration, *Digital Alienation* (see *Chapter 5. Digital Alienation - ICT's as Socio-Technical Systems, p. 91*). However, it is used throughout the research project.

Our own investigations revealed that, though rarely recognised in law or politics, there are many different definitions of autonomy extant within philosophy. We organised them into six categories, such as procedural (thinking in certain ways), coherentist (acting in ways consistent with your values), and theological (acting as God intended). Rather than seeking to adopt a single definition of autonomy, we cross-referenced each category with each of ETICA's eleven technology groups so as to identify how each technology could restrict each version of autonomy. This cross-referencing used the premises listed in the preceding paragraph to understand the mechanisms by which each technology impacted human autonomy. This formed the first answer to the research question and was published as *Threats to Autonomy from Emerging ICTs* (Dainow 2017b)[2].

1.4.2. Phase Two – accounting for the impact of multiple ICT's on multiple versions of autonomy

However, while this paper adequately answered the research question in terms of the typical approach in technology studies, and would therefore be considered a satisfactory answer by many, we felt alternative approaches could be valuable. While ICT ethics typically explores technologies in isolation, sociological and psychological research reveals that humans do not respond individually to each technology as if they were unconnected, but to the combined effects generated by all of them impacting at the same time. Furthermore, these findings accord with the intended aim of ICT innovators to create a seamless integration of personalisation services such that the switch from one delivery system to another as the person moves about is invisible to the user. We first

[2] The methodology therein treated the social dimensions of technology as the same across all ICT's. We recognise that there may be individual differences with specific contexts, but we also recognise there are many common factors. Human autonomy, no matter how conceived, involves a long-term element of one's life, and, under all accounts, involves the relationship between current actions and long-term dispositions or values. While there may be important variations in social dimensions of ICT's within different contexts, and these may have ethical dimensions, we believe there is sufficient commonality of effect, mode of interaction and social dynamics that one can understand the impact on human autonomy of any individual ICT through shared social mechanisms. However, any substantive argument to prove this would take us too deeply into the philosophical analysis of autonomy and the sociology and philosophy of technology and would therfore be beyond the scope of this research.

encountered this when reviewing statements made by leading automotive innovators at the TU Automotive Detroit 2016 Conference (Bedigian 2016), such as

> "We very quickly discovered that customer expectations really aren't based on our Jaguar Land Rover competition – they're based on the smartphone. They're based on the tablet. They're based on the home entertainment experience, which kind of makes complete sense." - Matt Jones, Director of Future Technology, Jaguar Land Rover (Bedigian 2016, 7).
>
> "[Consumers] want their seamless life. It doesn't matter if we're here in North America if we're in China or if we're in Japan or India, that seamless life is where people want to be." - John Schnoes, Programme Director of Vehicle Information Technology and Autonomous Drive, Nissan (Bedigian 2016, 7).

We are also aware that companies may attempt to own the entire ecosystem around any commercial service in order to dominate their market (Hannaford 2007) and that this trend is especially strong in ICT's (Fontana and Nesta 2009; Assink 2006). We therefore considered it appropriate to extend the research to consider the manner in which autonomy could be threatened by this combined effect of a seamless digital environment.

The methodological problem here was the lack of a pre-existing framework which could both describe this combined effect and account for the mechanisms by which it impacted autonomy. We therefore developed a formal methodology for describing and examining ethical processes within multi-dimensional technological environments. This was published as *Smart City Transcendent - Understanding the Smart City by Transcending Ontology* (Dainow 2017a). This framework was then used to develop descriptions of the mechanisms by which an all-encompassing digital environment can restrict human autonomy, found in the book chapter *Binding the Smart City Human-Digital System with Communicative Processes* (Dainow in-press). This latter publication also identifies systemic dangers present in all ICT innovation and operation which can threaten human autonomy, irrespective of the nature or purpose of that ICT. These latter two publications constitute a second alternative, but compatible, answer to the research question.

1.4.3. Summary – how did we answer the research question?

We have answered the research question twice. The third and fifth publications directly answer the research question, though in different ways. The third publication, *Threats to Autonomy from Emerging ICT's* (Dainow 2017b), takes the technologies identified by ETICA and explores each individually to determine if it can be a threat to any of the major versions of autonomy currently extant in modern philosophy. Where a threat exists, the paper identifies which versions of autonomy are threatened by that particular technology and how. However, this account treats each technology in isolation and does not account for their combined effect. Treatment of technologies in isolation like this is typical in ICT ethics, computer science and technology studies and so the answer provided is coherent with mainstream practice within these disciplines. The fourth publication, *Smart City Transcendent* (Dainow 2017a), develops Integrated Domain Theory as a way to analyse the combined effects of all technologies on the individual by treating them as a "digital environment." The fifth publication, *Binding the Smart City* (Dainow in-press), uses Integrated Domain Theory to provide detailed accounts of the processes by which this digital environment can threaten human autonomy. The advantage offered here is that Integrated Domain Theory can show restrictions on human autonomy which apply to any and all versions of autonomy.

The two answers are complementary. *Threats to Autonomy from Emerging ICT's* is of value to those wishing to explore the ethical impact of a specific technology. In particular, by surveying all major accounts of autonomy in context with each technology, it extends the more common analyses, which typically focus on a single version of autonomy (and usually fail to recognise that it is just one of many versions). *Binding the Smart City* will be of value to those who work at the level where multiple technologies are considered in combination, such as urban planning (especially smart cities), architecture, sociology and broader social analyses within computer science. Here the ethical concern is with populations of individuals, who will have differing conceptions of what it means to live autonomously, yet have equal right to consideration.

1.5. Contributions

This research offers a number of contributions:

1.5.1. Resolving difficulties in marxist analysis of the digital economy

A central tenant of marxist social analysis is the concept of alienation (Ollman 2001). Simply put, under this view, people are coerced into undertaking work in which they have no interest so that their labour may be exploited in order for their employers to extract a profit. The problem which confronts marxist analysis is that value is extracted online from voluntary activity. This is activity which people have not been coerced into undertaking and to which people feel connected. Furthermore, as this is not paid work, it does not fall under the traditional definition of labour. This generates difficulty for marxist analysis of the internet economy because of the presumption that value can only be extracted from paid labour and that no one would undertake such labour unless coerced into doing so. The approach to these problems within the scholarship has been to attempt to cram the internet economy into the traditional terminology, usually by redefining key terms, such as 'labour.' For example, it has been argued that labour is not paid employment, but any activity which someone else can gain value from (Scholz 2013). *Digital Alienation as the Foundation of Online Privacy Concerns* (Dainow 2015a) bypasses these difficulties by instead retaining Marx's original definitions and accepting the traditional marxist analysis regarding how value is derived online, the nature of coercion, and so forth. However, this accepted state of affairs is used to argue that Marx's mechanisms by which alienation occurs in the traditional workplace are not applicable to the internet. Instead, the paper offers an alternative mechanism specific to the internet economy, under the term '*digital alienation*.' In other words, we retain the terminology and instead update the processes by which alienation arises. Digital alienation occurs where digital systems impose on the user modes of life which are contrary to their wishes or which fail to allow them to express aspects of themselves or their life in the manner which they would wish. The paper explains the mechanisms by which this occurs, taking into account both engineering aspects, such as the variables used to encode information about people, and social aspects, such as business model, as well as specific models of perception and cognition. While the description of this mechanism is specific to the modern internet economy, it does so in a generic manner which is also applicable to the future state of affairs our research is concerned with. The changes to Marx's account are intentionally as small as possible so as to retain the majority of marxist analysis. Given the presence of ubiquitous commercial surveillance as the originating source of income on the internet, we simply replace labour with existence. In other words, people are exploited on the internet simply by using it, not by engaging in specific forms of activity. Not only does this resolve the difficulties marxist scholars have been struggling with, it better positions marxist analysis for the future state of affairs in which digital systems, such as IoT, will record all human activity (and possibly speech) and value will be extracted from this data by service providers. This contribution is of value in media studies, social analysis within computer science, critical theory and any other analysis of digital environments working from a marxist perspective. Simply put, the digital economy does not exploit labour, but existence itself. We do not believe marxism will be capable of understanding the smart city and other future digital environments unless it follows this move.

1.5.2. Development of a formal, generic, methodology for evaluation of large foresight research projects which combine multiple methodologies.

This evaluation protocol is not dependent upon the individual methodologies used in a foresight studies research project. The lack of such a methodology, and the need for it, is well documented in foresight studies. We believe, as do many others, that a formal methodology is needed to ensure all important aspects of a research project are evaluated. While there are accepted evaluation methodologies within individual disciplines, research projects like ETICA's combine multiple methodologies. It is not sufficient to evaluate these methodologies individually since it may be the way in which they were combined which undermines the research findings. Indeed, we found this to be the primary criticism which could be made against the ETICA project. It is therefore necessary to evaluate such multi-disciplinary research at a level which considers the coordination of multiple disciplines. For example, one concern is to determine whether the data output from one step in the research process was accurately passed to the next when that next step uses different methodologies and datatypes, or whether it was "translated" in a manner which lost essential elements. This need is recognised in the field and several attempts to create such evaluation techniques were considered. The most widely used assessment protocol is Impact Assessment (Poteralska and Sacio-Szymańska 2014). However, this does not evaluate the methodologies or validity of the findings, but the changes in planning and policy which result from the findings. Even here, no formal methodology exists, so much as non-specific approaches. Medicine frequently encounters the same issues with multi-disciplinary research and has developed a formal methodology, CASP (Singh 2013), for evaluation. The chapter evaluating ETICA's technology findings which we use as our dataset offers arguments in favour of adapting CASP to foresight studies (see *3. Evaluation of ETICA's Technology Analysis, p.33*). We then develop a formal CASP evaluation technique from multiple versions of the CASP methodology. The subsequent CASP evaluation of ETICA reveals variance between researchers as to what the aim of the project was. This resulted in variations between the project's reports. However, these were not so severe as to undermine those of ETICA's findings which were used in this research project. We believe this adaptation of the CASP approach offers the most detailed and structured methodology for evaluation of foresight studies research extant today and is therefore of significant value to the discipline.

1.5.3. A catalogue of potential threats to human autonomy across the range of emergent technologies.

Of particular import is the recognition that there are competing versions of autonomy and that the catalogue of threats published accommodates all of them. This was published in *Threats to Autonomy from Emerging ICT's* (Dainow 2017b). This is the first time any analysis of autonomy under ICT's has considered the different versions of autonomy rather than assuming there is just one (which is rarely defined). The paper identifies six species of autonomy within modern philosophy. While some versions simply disagree on relevant factors, others are diametrical opposites. This is particularly important for any area considering issues of consent, including design of systems for assisted living, bioethics and legal development. It renders attempts to predict a new technology's impact on autonomy much more complex, but also allows for anticipation of aspects only a minority may object to. For example, under some versions of autonomy it is possible to argue that it violates human rights to be educated or nursed by a robot. It also explains why it is possible for some to view a technology as liberating, while others view exactly the same thing as restricting their autonomy. It cannot be overstated how important this is nor how rarely it is mentioned in the relevant literature. Even when it is acknowledged that more than one definition exists, it is typically assumed that the concept can be treated at a higher level without reference to internal details, such as by considering context, aims or outcomes. However, this fails to recognise that even such factors as these may not hold consistent across all definitions of autonomy. This contribution is of value to any ethical analysis of ICT's which is concerned with human rights or any other rights or values dependent upon autonomy. This includes future efforts regarding legislation or other efforts to maintain or enhance human autonomy.

1.5.4. Development of Integrated Domain Theory, a detailed conceptual model by which to formalise accounts of the ways in which multiple ICT's ethically interact with the individual.

Integrated Domain Theory is the focus of the latter two publications and comes out of our research into the smart city. Having determined that humans are affected by ICT's in combination, rather than as a set of discrete individual technologies, we sought for a framework by which to describe and assess the future of emerging ICT's without differentiating between them. This framework was found in the concept (or vision) of the smart city. However, ICT ethical literature on the smart city and most other computer science literature on the smart city, focuses on individual technologies, rather than the city as a whole. Where the city as a whole is discussed, there is no linkage with the individual technologies of which it is composed. That this is a difficulty is indicated by the number of attempts to construct a model which can accommodate both top-level management systems and lower level devices, a survey of which is provided in our fourth publication, *Smart City Transcendent* (Dainow 2017a). Most models within ICT ethical literature (and computer science literature in general) follow the template of layered models similar to the Open Systems Interconnection (OSI) model (OSI Project 1994). Urban Planning, Architecture, Geography and other disciplines which deal with the built environment have a number of conceptual models and supporting methodologies by which to consider the urban environment and its inhabitants as a single system. These are necessary in order to cross-reference between macro-level aspects, such as street design or urban management, and the effect on people, and vice versa. A popular approach within these fields derived from sociology is Actor Network Theory (ANT), which is more methodology than theory (Latour 2005). The essence of ANT involves treating the subject as a system and tracing patterns of interaction, especially cause and effect, between nodes while ignoring what the node is made of. Thus, it is possible to consider the impact on human behaviour of an inanimate sign, treating the sign as a cause of behaviour. Similarly, it is possible to treat the sign as being caused by human behaviour. ANT treats all technology as socially embedded through mechanisms of non-linear causality. Our research into smart city literature within ICT ethics revealed a lack of awareness of ANT and its potential to resolve analytic difficulties encountered when trying to deal with the complexity of the smart city's ICT infrastructure and its interactions with people. However, we did not consider ANT alone sufficient for a computer science analysis because it lacks components to describe the unique properties digital devices possess which render them able to create a cohesive digital environment. In that societies are organised from within by their members, they can be said to be self-organising, or *autopoietic*. Autopoietic systems have been examined in both biology and sociology, but also exist in IT systems such as TCP/IP's dynamic routing (Graziani and Johnson 2008). We fused the sociological work on autopoietic social phenomena with ANT to produce a descriptive framework (with an implied analytic methodology) for any complex environment composed of both digital systems and humans. The first portion of this work is found in the paper *Smart City Transcendent*, while a second portion is found in the book chapter, *Binding the Smart City*. This latter work goes into more detail regarding the nature of the nodes and uses the framework to explain the different mechanisms by which ICT's may restrict human autonomy. This is primarily of value in the social dimensions of computer science, especially with regard to issues of algorithmic justice and automated decision making, software modelling, smart cities and other "coded" environments. It is the only methodology available within ICT ethics by which to account for complex socio-technical relationships, where we may want to trace relationships back and forth between human activity and engineering feature, or between across ontological levels, such as tracing the relationship between an API standard and automotive traffic patterns. The value of this methodology is exhibited in *Binding the Smart City*, which demonstrates how the model can be used to identify specific autonomy-limiting characteristics applicable to any ICT and any context. ANT has already proven its value in other fields. The contribution here is to bring ANT into computer science by adding new features meeting the requirements unique to the discipline.

1.6. Findings

This section briefly outlines the overall findings of the research process by summarising the accounts of Integrated Domain Theory found in the fourth and fifth publications.

Just as we cannot separate the impact of one technology from another when looking at how they affect the human, so we cannot remove the human element from the technology itself. If we are to examine human autonomy, the phenomenon we must consider is an open system composed of all the technologies which impact on a person and the mechanisms by which the system can restrict autonomy. This must include the human element which designs and delivers the system's components and largely determines the uses to which they are put. Modern society is moving into a state of affairs which can best be characterised as a new form of environment. Until now, humans have only had two types of environment to inhabit - the natural world and the built urban environment. We are in the process of creating a third type of environment which has not existed before, characterised by increased intelligence and response capabilities. This "digital environment" has such a deep and wide-ranging integration with human society and the individual's personal experience that it constitutes a new way of living. Our fourth publication, *Smart City Transcendent*, argues this can be best described as a self-aware autopoietic system composed of human and digital agents equally, which we term an *"Integrated Domain"*. Here the human and the digital are integrated to such a degree that processes or events are not caused by either a human or a digital process, but by the two acting in a manner which makes them inseparable. Ways of living will emerge within this hyper-complex system which cannot be anticipated now, but will still exhibit the power dynamics we see in current society, in which organisations use technologies to vie for economic and political power in a pluralistic society. On the basis that ICT technologies have not fundamentally changed the nature of society so far (Curran, Fenton, and Freedman 2012), we do not anticipate future technologies are going to do away with the tensions between individual and institution we have experienced since the rise of complex societies 10,000 years ago.

Threats to autonomy will occur wherever the human user would want control over those aspects of their digital environment which they consider necessary for their autonomy (however they conceive it) which are not available to them, but which could have been available to them had some aspect of the technology been different. Our first two publications demonstrated there are many existing trends which could lead to a lack of autonomy. The worst case scenario for the final outcome of these trends can be described in terms of previous social structures which exhibited similar restrictions on autonomy in the past. These are feudalism, totalitarianism and slavery. Current trends in innovation have the capability to lead to a situation in which one's ability to be alive becomes dependent upon renting services under terms of obligation from an elite, resistance to which results in (at minimum) social deprivation or (at worst) death; a feudal system which permeates, surveils and controls every aspect of one's existence - the ultimate totalitarian vision.

In sum, the threat to autonomy is that the digital environment into which we are emerging will be controlled and used by a few in a manner which reduces the autonomy of the many, or which denies them new liberatory forms of autonomy which the technology could have produced. At its worst, this new environment may create levels of oppression unseen since the Middle Ages.

2. Background

This chapter contextualises the research project. We will briefly introduce the topic, the main disciplines used and how they relate to the research questions. We will also provide a short summary of the ETICA foresight studies project, which provided the initial data for analysis.

2.1. Disciplinary Context – Computer Science

This research falls into the discipline of computer science. There are various ways of defining the field of computer science, such as whether it is a branch of engineering or science (or both), or whether the primary foundation is the study of algorithms or information (or both) (Rapaport 2017). However, all but the narrowest definitions include a role for consideration of the context in which the discipline exists, including its relation to other disciplines. Much significant work in computer science involves collaboration with other disciplines, either through drawing on research, concepts or methodologies, or through collaborative research projects. Computer science overlaps with philosophy in two respects; as a topic of philosophical enquiry itself and as a branch of applied philosophy. As a direct topic of philosophical enquiry, common concerns are the nature of the topic, such as the relationship between data and information or the relationship between an algorithm and code, and the nature of the discipline, such as what form of science it is. Applied philosophy is concerned with the use of philosophy within other disciplines. The central methodology is the application of traditional philosophical methods, concepts and concerns to a particular discipline or to the context of that discipline. Computer science can, under most definitions, include the study of the use and consequences of the products of its discipline (Schneider and Gersting 2010; Rapaport 2017). For example, *Invitation to Computer Science* (Schneider and Gersting 2010) uses a seven-layer model to organise the field of computer science, consisting of an ascending ontological scale of algorithms, hardware, abstraction layer (their term is 'virtual machine'), software, applications and social issues. Social issues of computer science have been most frequently examined using techniques of sociology and psychology under the term 'technology studies' (Brey 2005). A relatively new, but growing, field is the application of ethics to computer science under the often misleading title of 'ICT ethics'. ICT ethics may be defined as the branch of computer science concerned with the ethical dimensions of the topics studied within computer science (Spiekermann 2015). This can include the ethical dimensions of methodologies and procedures as well as the application of research in technology innovation (and the consequences thereof). Because ethical positions frequently depend on underlying positions in other philosophical fields, including metaphysics, epistemology, philosophy of language, philosophy of mind, political philosophy and philosophy of law, these fields may be drawn into ICT ethics and applied to computer science. This research project is concerned with ethical aspects of the potential consequences of current computer science activity. For reasons listed above, the ethical dimension is limited to human autonomy, while the potential consequences are limited to threats to, or limitations upon, that autonomy. In that this research looks to the future, it combines ICT ethics with foresight studies.

2.2. Disciplinary Context – Foresight Studies

This is a work of intentionally negative imagination. This research asks "how bad could it get?" The objective is to develop a description of the ways in which future ICT's could make things worse. As a philosophical work, this research distinguishes itself from most foresight studies, which are more usually concerned with extrapolating a future state of affairs from current trends (Glenn 2009b). We do not seek to explain how current states of affairs will lead to the future digital environment, but to evaluate the place of human autonomy within that future state of affairs. We do not seek to assess the probability that such a future state of affairs will occur. For reasons which will be explained below (see *2.4 Dataset: The ETICA Project, p.30*), we do not seek to explain the mechanisms by which the current state of affairs could lead to such a future. Our role is simply to understand human autonomy and its interaction with ICT's to the degree that we can postulate the mechanisms by which autonomy could be threatened by ICT's in the future. Doing so requires a detailed understanding of autonomy, which is a much more multifaceted concept than most recognise when they use it. Similarly, emerging ICTs and their direction of travel must be understood in detail. A significant methodological problem occurs in the unpredictability of future technologies and how people will use them. The solution requires a way of talking about human-ICT interactions in the future which can discuss ethical issues in detail without going into specifics about technical operations. Such a methodology does not exist and so had to be developed as part of the research process.

'Foresight studies' refers to a variety of methodologies by which to undertake a structured examination of the future. Ways of describing the *current* state of the world are organised into the disciplines. Foresight studies uses those same systems for describing *future* states of affairs. Foresight studies must therefore overlap with the discipline which provides the desired way of looking at the world and may combine several such disciplines (Glenn and Gordon 2009). As a work of applied philosophy, our overarching approach is exploration of ethical concepts in a philosophical, not empirical, legal or technical mode of enquiry. Ethical implications of emerging ICT's are contextualised within the philosophical tradition, though some reference may be given to their treatment in other disciplines. In that this work must account for the mechanism by which ICT's can affect human autonomy, which is a property of individuals, it must include an account of the manner in which technical features create psychological effects which is at least coherent with psychological theories and empirical evidence. In that it examines socio-technical systems, it must include some sociology. Our primary focus here is those aspects of the social which stand between the technical artefact and the psychology of the individual. As we demonstrate in the fifth publication, *Binding the Smart City* (Dainow in-press), it is here that mechanisms which promote or inhibit autonomy are to be found.

We are not concerned with the internal details of technologies unless relevant to understanding their ethical impact. Our concern is for the effect a technology has on people, not the manner in which it works. As a system, technology is subject to common axioms of General Systems Theory, most importantly *equifinality*. In an open system the same final outcome can be achieved in different ways (Von Bertalanffy 1968). Given that any specific output can be achieved by more than one method of technical construction, this means we consider *types* of technologies rather than individual instantiations. The engineering details become relevant only when we consider whether a given technology *must* have a given ethical effect or whether it is technically feasible to create something which achieves the same purpose without that negative consequence.

2.3. Research Approach

2.3.1. Premises

The purpose of this section is to demonstrate a systematic analysis of the validity of research question, with particular focus on whether the question has been framed in a manner which assumes debatable premises, disagreement with which could undermine the project's findings.

The primary research question raises some secondary questions which could legitimately have been investigated within the research topic, but which were not. Instead, answers to these questions were accepted as premises. A position was accepted as a premise where there existed a general consensus in favour of it, such that acceptance would generally be considered unremarkable or reasonable. Potential questions were so treated where the scale of the task required to investigate the matter would have compromised research into the primary question due to time constraints.

The secondary questions "what are the emerging ICT's?" and "what is human autonomy?" address the meaning of key terms in the primary question. Another question intrinsic within the primary research question is the connection between these terms; i.e.: "is it possible for an ICT to threaten autonomy?" This question considers the role of technology within society and what, if any, impact technological changes can have on the society in which it occurs. This is a complex question, but the general consensus is that technological change causes social change, though there may be disagreement regarding how much and in what way (Brey 2005; Disco 2005). Mainly because of the scope of the question, but also considering the broader consensus, this research project does not investigate this matter. Instead, it is accepted as a premise that technological change causes social change and that this social change occurs though the aggregated effects of technological features upon individual people. Similarly, we accept as a premise that social factors significantly influence the nature of technological innovation.

For similar reasons, we also predicate this research on the premise that society can be analysed as a system (Parsons 1991; Luhmann 1995; Von Bertalanffy 1968; Latour 2011), and therefore may exhibit emergent properties not discernible from the properties of the individual nodes. We also accept that social changes are mediated through psychological changes in the individual and so social change cannot occur unless individual's change their behaviour (Castoriadis and Blamey 1997). This is not a reductionist view of society as nothing more than aggregated individual behaviour. In accord with our acceptance of society as a system, we accept the premise that changes at an individual level may result in the emergence of features at a social level not predictable from the behaviour or other properties of individuals.

2.3.1.1. Premises in ICT innovation

The research accepts some common premises about trends in ICT innovation[3]:

Miniaturization

Moore's Law has been a reliable summary of key trends in ICT innovation since the 1960's (Powell 2008). The number of processing units which can fit into a given amount of space has been increasing continually for many decades (Mollick 2006). While there are some concerns this may be slowing, others argue it can be sustained through strategies like improved coding (Palem et al. 2009) and better energy supply (Dreslinski et al. 2010). While Moore's Law refers to the reduction in size of circuitry, it also reflects the resultant reduction in size of the devices housing them. We expect computing devices to continue to reduce in size and it is possible nanoscale devices will be common within the timeframes under consideration, though our findings are not dependent upon this. We also note the development of smart materials (Addington and Schodek 2005), which may make the distinction between computing device and the material in which it is embedded obsolete, or even provide "smart" capabilities to traditional materials, such as wood (Ugolev 2014; H. C. Kim et al. 2016).

Convergence

Miniaturization allows for more functions to be embedded within a device of any given size. Over time, ICT's have "absorbed" capabilities of other devices (Mueller 1999). For example, the smart phone has absorbed the capabilities of the telephone, the still camera, the movie camera, the voice recorder, the fax and the radio. In addition, there is a tendency in the ICT industry for larger companies to absorb smaller ones, both as a competitive strategy and as a means of acquiring innovations (Gates 2000; Hacklin, Marxt, and Fahrni 2009). We therefore assume both forms of convergence will continue; that devices will become more powerful and multi-functional and that the industry will, unless prevented in some way, contract to near monopoly levels.

Intelligence

Miniaturization combines with ongoing developments in computer science such that devices become more "intelligent." Here "intelligence" is founded on increasing rates of instructions per second, which make possible more complex

[3] While these are commonly accepted, it is worth noting that ETICA reported finding empirical evidence of these trends in their raw data, though this evidence was not published.

processing in any given unit of time, in combination with development of more complex algorithmic processes, such as those seen in machine learning and neural networks. Over time this leads to the ability of ICT's to take into consideration more variables and to do so in a wider variety of ways. We expect this trend to continue, especially in the development of soft AI systems. We do not anticipate the rise of hard AI systems which emulate human consciousness within the timeframe considered within this study. Indeed, we are sceptical that such emulation is possible at all. However, we do anticipate that future ICT systems will be able to process context and social nuances to a degree not possible today.

Ubiquity

The above trends lead to wider usages for ICT's, an improved price-performance ratio (Barroso 2005), and a wider range of environments in which ICT's can be placed, making any given usage progressively cheaper and easier. As a result, ICT devices have spread throughout society and across the globe (M. I. Wilson and Corey 2011). We expect this trend to continue. We anticipate that it will be practical to install ICT systems in places and for roles which are currently uneconomic. We therefore anticipate that the future urban environment will contain many more ICT's – to the degree that the absence of some form of ICT in any given context will become notable (Dodge and Kitchin 2005).

2.3.2. Normativity & Feasibility

This research is a normatively neutral ethical study. It examines an ethical issue but does not recommend any responses to that issue. The goal of this research is to determine how human autonomy can be threatened. The individual reader will decide for themselves the role and value of human autonomy. From this they will determine the severity, or otherwise, of the threats identified in this research. It follows that it is not the task of this research to recommend remedies to avoid such dangers. Remedies are based on one's assessment of the importance of the threats identified, especially when it comes to trade-offs against other benefits. However, as will be discussed later, the treatment of autonomy herein is often founded on the availability of choice. We do not consider autonomy is restricted if something is physically impossible. It follows that an ICT's operations do not restrict autonomy where alternatives are technically impossible. A key point argued for repeatedly in our publications is that many specific implementations of tasks in ICT's are treated as unavoidable on the mistaken assumption there is no technically viable alternative. It is therefore frequently necessary to demonstrate the technical feasibility of alternative ways of achieving the same ends which do not restrict autonomy. Offering technical alternatives is therefore a side-effect, not goal, of this research. As with assessment of severity, the advisability of preferencing such alternatives will depend on one's assessment of the role and value of autonomy in human flourishing. It would be a mistake to believe this normatively neutral position means we assume any "reasonable" person would take a position in favour of human autonomy and oppose features of ICT's which restrict it. This is not the case. Our publications make several references to the Chinese Social Credit system, a vast complex of ICT's in deployment now which will eventually permeate most aspects of the Chinese citizen's life, and which is designed specifically to restrict individual autonomy and remove "social contradictions" (State Council of People's Republic of China 2014, 3). Though it is outside the scope of this work to offer detailed examinations of cultural differences towards ethical values, we do not accept that emerging ICT's will be governed by the same values in all parts of the world. Instead, we assume that different countries will evolve in different ways, such that we will see a range in the ways autonomy is restricted.

Similarly, this work seeks to assess ICT innovation against the widest possible range of accounts of what constitutes autonomy, and has not taken a position in favour of one account. Indeed, the position taken here is that thinking there is a single correct account of what constitutes human autonomy is to misunderstand the concept itself. Instead, we examined each definition of autonomy individually to determine how that version of autonomy could be threatened. By cross-referencing each form of emerging ICT with each form of autonomy, we were able to deduce features any ICT must have if it is to restrict autonomy under all definitions. These are defined in the third publication, *Threats to Autonomy from Emerging ICT's* (Dainow 2017b). Use of this understanding is demonstrated in two subsequent publications, *Smart City Transcendent* (Dainow 2017a) and *Binding the Smart City* (Dainow in-press). The intent is that, by remaining at the level of discussing all forms of ICT's and all forms of autonomy, this research is relevant to the widest possible community.

2.3.3. Our use of foresight studies

2.3.3.1. Background

'Foresight studies' is one of the terms used to describe the discipline of considering the future in a methodical manner. Foresight studies are not attempts to predict the future, but to reduce uncertainty about it (W. Bell 1996). Foresight studies can be used in any discipline and so the objectives in such research vary widely. What will be considered depends on who is considering the future and why (Glenn 2009b; Godet 2009; Veikko, Kanerva, and Kouri 2009). Typical examples include assessment of existing visions of the future, the implications of new technologies, identification of opportunities or threats in the future and the assessment of possible impacts of policy decisions (Popper 2009). This growth within many disciplines has led to a variety of terms being used to describe the activity. Common alternative terms for foresight studies are "future studies" (most commonly used in the USA), "futures research" (Europe), "futurology" (Australasia), "prospective studies" (France), "futures field" (Europe) and "prognostics" (Russia) (Veikko, Kanerva, and Kouri 2009).

The origins of foresight studies are typically traced back to H.G. Wells' 1902 widely-disseminated lecture at the Royal Institution, *The Discovery of the Future*, in which Wells described the possibility that science could produce "an ordered picture of the future" (Wells 1913, p.36). Future-looking policy research commenced formally in the 1930's under the US government, becoming an independent discipline with the formation of the RAND think tank in 1946. The 1960's saw the rise of non-commercial foresight studies with the founding of institutes such as Bertrand de Jouvenel's Association Internationale de Futuribles (1960), Herman Kahn's Hudson Institute (1960) and the American Academy of Art's Commission on the Year 2000 (1965). The field can be said to have reached maturity with the publication of Wendell Bell's *The Foundations of Future Studies* (W. Bell 1996).

2.3.3.2. Our approach to ethical foresight studies

A wide variety of foresight methodologies have evolved and most foresight research projects combine several methodologies. A survey of 1,000 foresight research projects across sixty countries identified twenty-five discrete methodologies being used (Popper 2009). Conducting their research over the same time period, Glenn and Gordon identified thirty-eight different methodologies used in foresight research (Glenn and Gordon 2009). It is generally accepted that a number of foundational premises underpin all foresight research. Foresight studies are epistemologically premised on the position that there is an objective reality which exists independent of our knowledge of it. Foresight researchers generally agree that the future is not predetermined and that people can influence the development of the future through their own actions (Aaltonen and Barth 2005; Veikko, Kanerva, and Kouri 2009).

Foresight research is not, therefore, an attempt to predict the future, but to develop "conceivable" future states of affairs. It is expected that some of the elements described will occur in some form, but it is considered unlikely in the extreme that the future will exactly match the state of affairs described in the research project (Glenn 2009b). One of the field's pioneers, Herman Kahn, argued that history shows that unpredictable events frequently cause radical changes and that much of history would be completely implausible, if it were not for the fact it actually had happened (Kahn and Wiener 1967). A common aim is to identify possibilities for change which could influence key aspects of potential futures. Perhaps the most well-known example of such a "foresight effect" is Orwell's "*1984*" (Orwell 1983). While no one expects the world it describes to come about, the concept of pervasive government surveillance encapsulated under the label "Big Brother" has served as a powerful symbol with real political impact (Kellner 1990).

This approach is central to our own methodology. At all times we have sought to avoid analyses which are based on predicting specific technological implementations or human practices in the future. Instead, we have confined ourselves to consideration of classes of functionality, especially where these can be described in terms of their interactions with people. In Aristotelian terms, we have focused on teleological analysis (final causes) and processes (efficient causes) (Hocutt 1974; Aristotle 2014b). Our concern has been with higher level functional analysis, rather than the form in which technologies are instantiated in specific instances.

The different methodologies used within foresight studies can be divided between quantitative and qualitative in technique, and between normative and exploratory in purpose, though some methodologies bridge these divides (Veikko, Kanerva, and Kouri 2009; Haegeman et al. 2013; Gordon and Glenn 2004). Ethical assessment of the future is rarely a formal component of foresight research projects. This lack of ethical assessment in foresight research is widely recognised (W. Bell 1996; Poli 2011; Brey 2012; Clarke 2005; Grunwald 1999; Harris et al. 2010). It has been argued the dominance of the paradigm of the material sciences within foresight methodologies promotes a belief in "universal" ethical values, which, like the constants of physics, are applicable in all places at all times (Poli 2011). Philosophy has pursued a quest to discover such universal ethical values for literally thousands of years without achieving any consensus (MacIntyre 1998). Key foresight research figures like Wendell Bell have attempted to determine what ethical values are universally accepted by all peoples in all cultures (W. Bell 1996). However, Bell has since been criticised for accepting as objective and universal what are merely western cultural values (Poli 2011; Dator 2011; Rubin 2011). A less controversial approach has been to incorporate ethical values as research data. For example, the functionality of future technologies can be drawn the visions of those creating them, after which philosophical sources can be mined for ethical treatments of these functions. This was the approach taken by the ETICA project and our own work. It is for this reason we do not ground our normative assessments in either marxist or capitalist premises, but seek to view the same subject matter from both. We regard our assessments as strengthened wherever both forms of analysis come to the same conclusion, though they may use different terminologies to describe the same phenomena. It is also for this reason that we do not preference one version of autonomy over any other, but seek to examine all major accounts.

2.4. Dataset: The ETICA Project

The ETICA project was a large foresight studies research project designed to identify emerging ICT's and outline their ethical concerns. Our research project used ETICA's technology groups as the emerging ICT's for analysis. We did not examine any individual technologies which ETICA had used to develop this set, nor did we seek to validate the assignment of an individual technology to a particular group. Much of ETICA's original data is no longer available, so it is not possible to reproduce the ETICA research leading to the ICT taxonomy. Validating it would therefore have been an original research project of similar scale and scope to the original ETICA project, something beyond what could be accomplished within our own project. This acceptance of ETICA's technology findings was not reflexive, but based on a formal and detailed analysis of the technology identification phase, contained in the following chapter (see *Evaluation of ETICA's Technology Analysis, p.33*). This evaluation identified some issues within the research project which allowed for drift between individual researchers and for minor variations in normative stances between different reports. However, these were not sufficient to invalidate the technology findings. In addition, we monitored the developments in ICTs during the five years of our own research project in order to detect any ICT's which emerged since the ETICA reports were published. The only significant technology not anticipated in ETICA was self-driving vehicles. However, we determined that self-driving vehicles fit within ETICA's definition of robotics as "machines with motor function that are able to perceive their environment and operate autonomously" (Ikonen et al. 2010, 114).

Some use was made of ethical implications ETICA identified which pertained to human autonomy. However, ETICA did not attempt to undertake detailed ethical analyses. Instead, we used ETICA to identify topics of concern, which were then researched in peer-reviewed publications. ETICA's final recommendations, being intended for EU authorities, and concentrating on the need and methods for further research, did not form part of our own research.

The most significant issue found with the ETICA research was a preponderance of focus on technical characteristics, which contrasted with very limited analysis of the social aspects of technologies. This indicated the need to provide a more detailed account of the social aspects of technologies as socio-technical systems. This account forms significant portions of three of the five publications for evaluation; *Digital Alienation* (Dainow 2015a), *Smart City Transcendent* (Dainow 2017a) and *Binding the Smart City* (Dainow in-press).

2.5. Addressing the social aspects of technology

As indicated, the research to develop a social framework arose from the conclusion that ETICA's technology definitions focused primarily on functional specifications, with little reference to social dimensions. It was therefore necessary to develop an account of the social aspects of emerging ICT's to complement ETICA's focus on engineering. It was determined that the most advantageous strategy was to provide an account of the social which could be applied to all the technology categories, rather than attempt different accounts for each technology. There were two reasons guiding this decision. Firstly, doing so limited the scope of the research to what was feasible. It was considered that a social taxonomy of similar depth to the technical definitions in ETICA would require a research project large enough to be a Ph.D. in its own account. Secondly, by having an account of the social which spans all technologies, all technologies can be subjected to the same ethical analysis. Doing so places their ethical status on the same basis and makes possible the deduction of general principles applicable to all.

As a result, it was determined that before analysis of ethical impacts could be considered, it was necessary to produce an explanatory framework which could account for the mechanisms by which technical artefacts produce ethical effects. This research commenced with the formulation of the third secondary question: "what is the mechanism by which a technical feature can produce a psychological factor?" This was answered in the paper, *Digital Alienation as the Foundation of Online Privacy Concerns* (Dainow 2015a).

3. Evaluation of ETICA's Technology Analysis

This chapter seeks to validate ETICA's identification of what are the emerging ICT's with specific regard to the use of these findings as the primary dataset of emerging ICT's in our own research. The aim is to demonstrate that the technologies identified by ETICA were sufficient for our own purposes.

Our research design called for a pre-existing set of emerging ICT's we could examine. Due to the scale of such a task, it was beyond the scope of our project to undertake the research necessary to develop such a set ourselves. Furthermore, such work had already been done by the ETICA project, one of the largest technology foresight studies ever attempted in (Brey 2012). Our first task was therefore to critically assess the ETICA project's findings in order to determine their suitability as the foundational dataset of emerging ICT's. Our aim was to determine if ETICA's findings could be used as the basis for our ethical analysis, or if they needed enhancement, or were so inadequate that alternatives should be sought. It is important to note the limits of our assessment. Due to the scale of the task we could not assess ETICA's conclusions by seeing if we could reproduce them. Instead we focused on an analysis of ETICA's methodology. Our reasoning was that, if ETICA's methodologies (including their own data) were adequate for the task, the technologies ETICA identified were sufficient for our purposes.

We shall first provide background information regarding the project as a whole and then summarise the technology identification research process and findings. Finally, we shall develop a formal assessment methodology and offer a structured assessment of the ETICA research process[4].

3.1. About ETICA

3.1.1. Aim

The ETICA project's aim were to identify emerging information and communication technologies (ICT's) and the ethical issues these gave rise to and to make policy recommendations.

[4] This chapter contains argues for and develops a formal methodology for assessment of complex future studies research projects. However, as it does not address the research questions, no attempt has been made (as yet) to publish this work and so it does not constitute one of the publications for assessment. However, we consider the methodology outlined below of value to the foresight studies community and so anticipate it will be developed into a formal paper at some future stage.

3.1.2. Output

The project's ultimate goal was to produce a series of recommendations for those policy-makers and commercial interests whom the project had determined should be considering the ethical issues of ICT's as they emerged (Stahl 2011b). These are contained in the *ETICA Project Final Report* and in the project's governance recommendations, *D.4.2 Governance Recommendations* (Rainey and Goujon 2011). Summaries of recommendations were delivered to the European Parliament in 2011 and are contained in *Policy Brief of STOA on the ETICA Project* (European Parliament 2011). ETICA's findings have been fed into the work of the European Group on Ethics in Science and Emerging Technologies, part of the Bureau of European Policy Advisors (Stahl 2011b). ETICA was one of a number of related EU FP7 projects concerned with ethics, innovation and ICT's, such as the Governance for Responsible Innovation Project (GREAT) (http://www.great-project.eu/) and the Ethical Governance Of Emerging Technologies Project (EGAIS) (http://www.egais-project.eu/). ETICA contributors have produced papers in several journals, many book chapters and conference presentations.

3.2. Context

The ETICA project was an EU 7th Framework Programme (FP7) research project funded under the Science and Society funding stream from April 2009 to May 2011. The objectives of FP7 were to strengthen the EU's science and technology base, improve the EU's competitiveness and support policy development within the EU. ETICA's context within the EU superstructure was to support policy development. The ETICA project itself is an exercise in both ICT ethics and classical foresight studies.

3.3. ETICA's Wider Contribution

ETICA's primary impact has been with regard to methodologies by which to engage with ethical issues rather than on investigation of the issues themselves. The majority of ETICA's deliverable content is concerned with methodology, while the project's findings have mainly been used to develop further research *apropos* bringing ethics regarding ICT innovation into public discourse and policy development.

On March 31st, 2011, the ETICA consortium delivered its key findings to the European Parliament at a workshop delivered in conjunction with Science and Technology Options Assessment, Policy Department E: Legislative Coordination and

Conciliations, Directorate General for Internal Policies, European Parliament. The governance recommendations therein are exclusively concerned with mechanisms for engagement with ethical issues, not with the content of those issues themselves, such as, for example, recommending specific policies regarding specific issues. The main governance recommendations are aimed towards enhancing ethical discourse and ethical considerations within ICT development; development of a regulatory framework supporting (and possibly even demanding) ethical impact assessments for ICT research and development; the development of an ICT "ethics observatory" which would communicate the methodologies and content of ethical ICT discourse within academia; and the development of mechanisms for institutional discourse regarding ICT ethical issues (and research methodologies) within the broader range of civic institutions, such as NGO's and commercial enterprises (European Parliament 2011; Stahl 2011b).

ETICA's work was later incorporated into the research project *Framework for Responsible Research and Innovation in ICT*, funded by the UK Engineering and Physical Sciences Research Council, while the "ethics observatory" has since commenced in the form of the Responsible Research and Innovation Project (http://responsible-innovation.org.uk) ('ETICA Project Website' n.d.). ETICA's work is cited as foundational to the Governance for Responsible Innovation Project (GREAT) (Pellé and Reber 2013) and members of the ETICA project have continued work in related projects, such as EGAIS. Perhaps more importantly, ETICA's work has been cited in reviews of foresight ethics methodologies, such as *PHM-Ethics and ETICA: Complementary Approaches to Ethical Assessment* (Stahl, Mittelstat, and Fairweather 2013), *Anticipatory Ethics for Emerging Technologies* (Brey 2012) and GREAT's *DEL.2.2. Theoretical Landscape* (Pellé and Reber 2013).

3.3.1. The ETICA Project's Technology Identification

ETICA deployed a variety of established foresight studies methodologies, including bibliometric analysis (Porter 2009), expert groups (Glenn 2009a), scenarios (Glenn 2009c), scanning (Gordon and Glenn 2009) and morphological analysis (Ritchey 2009). Commencing with accounts of hundreds of ICT innovations, ETICA developed a taxonomy of eleven groups through identification of the essential functional characteristics which were shared by all members of that group. The focus for characteristics was identification of the manner in which they altered human being's interactions with the world (Stahl 2011b). ETICA found innovation paths in all ICT's shared some trends; convergence, increasing penetration of daily life, reusability and adaptability, reduction in size, increasing mediation of human perception and rising intelligence (Stahl 2011b). Note that many of these are the premises listed above as underpinning our own research (*see 2.3.1.1 Premises in ICT innovation, p.25*).

3.4. Ethical Issues Identified by ETICA

The most complete statement of the ethical concerns identified by ETICA can be found in *D.2.2 Normative Issues Report* (Heersmink, van den Hoven, and Timmermans 2010). Summaries of these issues may be found in other reports, such as *D.3.2. Evaluation Report* (Rader et al. 2010) and *D.4.2 Governance Recommendations* (Rainey and Goujon 2011). However, no other documents describe the ethical concerns to the same extent as the *Normative Issues Report* or offer any additional information regarding said concerns. Reportage of ETICA's ethical findings will therefore be drawn from the *Normative Issues Report.*

The ETICA project can be said to have identified ethical topic areas rather than specific issues themselves. No formal exposition of the issues was undertaken, such as statements of conflicting rights or duties, discussions of moral causation or consequence, applicability of universal moral principles or expositions of moral reasoning. For example; when discussing ethical issues associated with privacy and affective computing, the following analysis is offered:

> "For Affective Computing to be effective, a lot of personal data is needed sometimes even from external sources accessed via the web. Even when maximum efforts are taken to protect personal data and privacy we cannot go around the dreary fact that privacy will surely be eroded. All that remains is a recurring trade-off between the gains of affective computing and the loss of privacy."
>
> (Heersmink, van den Hoven, and Timmermans 2010, 23)

No further analysis is offered regarding this issue - the above is the complete text regarding privacy issues in affective computing. As can be seen, no analysis is undertaken to establish why the use of external data must inevitably lead to reductions in personal privacy or why the only solution to such a dilemma is to accept this. Furthermore, these projections are, to a large degree, based on two questionable but uninvestigated assumptions of ethical import regarding the manner in which such technology must be constructed; that additional data cannot be stored locally, but must be web-based and that algorithmic use of personal data must make that data accessible by third parties (and hence lead to reductions in privacy). Neither of these is a necessary aspect of the technology required to develop affective computing; they are merely extensions of current dominating, economically-determined models, of ICT development (Curran, Fenton, and Freedman 2012). There is no technical reason why it is impossible that all relevant data could be stored locally (Dainow 2015b; Langheinrich 2001). Nor is there any technical necessity that such data, even if stored elsewhere, should be accessible by others and therefore constitute a reduction in privacy.

Nevertheless, even if we accept these assumptions, it is clear that this passage does not constitute a detailed or developed ethical analysis. It does not, for example, detail what rights, duties, or other ethical values pertain to this conflict between affective computing and privacy, what moral or ethical precepts are involved, what relevant ethical theories or approaches are relevant and so forth[5]. As such, this cannot be described as an ethical analysis, but rather the identification of an area of ethical concern. A later section does discuss resolving these issues via informed user consent, but does not state the ethical values, processes or logic by which user consent would actually resolve such ethical issues. Furthermore, informed consent does not automatically or necessarily prevent a reduction in privacy, merely (at best) transfers responsibility or culpability for it by asking the user to agree to it. Because the discussion of reduced privacy does not elucidate the ethical variables involved in the matter, it is not possible for discussions of resolution via user consent to explain the logical operations by which this consent addresses or resolves the ethical concern.

Another example will suffice to show this level of analysis is consistent throughout the ETICA analysis. The section on neuroelectronics discusses the impact this area of technology has on the mind-body problem and related concerns:

> "It is also argued that Neuroscience demonstrates that mental states may be reduced to brain states. These reductionist concepts of the

[5] Ethical discussions within philosophy and ICT ethics may include a range of different elements. For example, discussion may include logical operations by which moral components of responsibility and obligation are transmitted via act and consequence or by mediation through ICT systems. There may, or may not be, reference to "universals" - values or schema held to apply to all people in all places. Some discussions may make reference to metaphysical or epistemological positions held to justify ethical stances. To some degree, the factors mentioned can be determined by the ethical school within which the discussion is situated. For example, Marxists may make reference to alienation while a capitalist neo-liberal is unlikely to recognise, or use, the concept...

> mind–body problem and the self entail serious ethical questions such as (a) Can the attribution of personal responsibility be reconciled with a neurobiological account of correlated brain processes? (b) Should we treat mental disorders merely as brain diseases? (c) Can the traditional notions of the unity and autonomy of the person be maintained in the face of their being questioned by neuroscience? Is the self only an illusion produced by the brain? Although the reductionist view of the self received counterarguments by several critics, for instance stating that the self cannot be found in somewhere in isolation, Neuroscience does affect the view of the self and consequently raises ethical questions."
>
> (Heersmink, van den Hoven, and Timmermans 2010, 83)

This passage demonstrates very clearly the level of analysis at which ETICA is working. The mind-body problem is one of the most complex and important problems in philosophy (Robinson 2011). Positions within the debate have strong implications for the nature of the human, law, politics, religion, civic and personal goods, and even the direction of science, to mention but a few. The above passage has not listed all the ethical issues associated with mind-body identification in neuroscience and could not do so without burdening the project's workload beyond feasibility. Given that several thousand years of philosophy has failed to resolve these issues, it is unreasonable to expect ETICA to do so. However, we can see from this passage that ETICA's analysis has merely identified a topic area and has not attempted to even list all the ethical questions raised, much less discuss them.

The above is not intended as evidence that the ETICA project failed, merely as evidence that the ETICA project did not attempt detailed ethical analysis. The aim of ETICA was to identify ethical concerns, not explore them in depth; much as an 18th century naval cartographer might map the location and outlines of a chain of islands, but not explore their interior. As such, it is not surprising that ETICA's recommendations focus on processes of engagement with ethical concerns, rather than the issues themselves.

3.5. ETICA Methodologies

> "The ETICA approach is a recent method for the ethical assessment of emerging information and communication technologies

> (ICTs). It is so general in scope, however, that nothing prevents its application to other types of technology as well." (Brey 2012, 5)

ETICA deployed a variety of established foresight studies methodologies, including bibliometric analysis (Porter 2009), expert groups (Glenn 2009a), scenarios (Glenn 2009c), scanning (Gordon and Glenn 2009) and morphological analysis (Ritchey 2009). Commencing with accounts of hundreds of ICT innovations, ETICA developed a taxonomy of eleven groups through identification of the essential functional characteristics which were shared by all members of that group. It developed formal structures for technological descriptions and scenario construction. Bibliometric analysis was then used to assess academic discourse regarding ethical issues associated with these ICT's. The project sought to validate and enhance these findings by using focus groups and surveys to assess corresponding concerns amongst experts outside academia. Finally, the project developed a series of recommendations, together with supporting philosophical and methodological expositions.

Our concern when evaluating ETICA was the methodology it used to identify the technologies of concern, so as to establish they are a reliable dataset for our own purposes We shall first summarise the processes ETICA followed to identify the emerging technologies, then undertake a structured assessment of that work.

3.5.1. Step 1: Development of data schemas

Identification of technologies and concerns was undertaken by "distillation of published views on these issues" (Stahl 2011b, 11). ETICA restricted project scope in the first instance by keeping its analysis at a general level, avoiding too much development of individual instances. This was accomplished through a focus on the essential functional capabilities of technologies, rather than individual artefacts. However, it was felt that reportage on this level alone could, in some cases, be too abstract to permit elucidation of specific ethical concerns on the grounds that ethical concerns are "always contextualised" (Stahl 2011b, 12). It was therefore decided to develop more detailed descriptions of example artefacts as illustrations for each technology, termed "vignettes" (Stahl 2011b, 12) . This represents an example of scenario construction (Stahl et al. 2010), an established foresight studies methodology developed by Herman Kahn in the 1950's and most famously pioneered by the RAND Corporation (Glenn 2009c). An additional rationale for the development of vignettes was that the range of possible ethical dilemmas was so large that it was "fundamentally impossible" (Stahl et al. 2010, 25) to identify them all[6]. It was determined that vignettes would offer the opportunity to provide illustrative examples, but it was also noted choice of vignettes could introduce researcher bias. It was hoped that bibliometric analysis would be uninfluenced by researcher bias and would compensate for this concern (Stahl et al. 2010; Brey 2012).

The process of identification of emerging ICT's by ETICA follows a classical Aristotelian approach to definition (Deslauriers 2007) through the development of an ontological taxonomy. The range of possible future ICT applications and the impossibility of describing all of them led the researchers to move the focus of attention to a higher ontological level and work at the level of the species rather than the individual instantiation. Each ICT species was characterised through the identification of the essential characteristics of that technology which would be shared by all instances thereof. The initial task required was therefore a determination of what constituted an ICT and what made it "emerging." Information and communication technology is a species of the genus technology, but what constitutes a technology is contended within the philosophy of technology (Introna 2011). It was determined that the defining characteristic pertaining to the genus of technology of relevance for the development of ICT species was that technologies alter the way humans interact with the world (Stahl 2011b)[7]. The defining characteristics of each ICT species would therefore distinguish the manner in which that technology altered human interaction with the world from the manner in which other technologies did so. Because such a "high-level system" (Ikonen et al. 2010, 3) could represent such a wide range of artefacts and situations it was further felt it needed to be "associated with a vision that embodies specific views of humans and their role in the world" (Ikonen et al. 2010, 4). A secondary criteria for the selection of defining characteristics was methodological - their potential utility towards the development of a "convincing normative issues matrix" (Heersmink, van den Hoven, and Timmermans 2010, 13). In addition to these two formally identified criteria, we

[6] From this we may understand that ETICA researchers consider "ethical concerns" to refer to specific instances of an ethical nature, not generalised ethical principles. Such instances are typically used within ethical discourse concerned with determination of appropriate actions, such as how one should act in such a situation or what governance framework is appropriate. Such instances may also be used as the basis for the induction of universals or to develop procedures for ethical reasoning. Given this framework, it would be unreasonable to expect ETICA to develop a list of ethical concerns in terms of universal principles or by offering generalised processes for resolution of ethical dilemmas.

[7] This position regarding ICT may be termed 'instrumentalist' and is far from universally accepted within philosophical, psychological and sociological discourse relating to technology (Introna 2011). This conception of ICT represents a particular and arguable position within philosophy of technology that must strongly determine the range of problems conceivable and their characteristics.

note the possibility that characteristics were chosen according to the degree to which they suited the normative positions of the researchers. Other ETICA documents, such as the *Heuristics and Methodologies Report* (Veikko, Kanerva, and Kouri 2009), state that all foresight studies evaluations should be, and cannot be other than, normative, rather than descriptive:

> "The practice of futures research has always a normative aspect to it because each scenario, forecast and image of the future have a potential and usually also an effort to affect the future. Therefore futurists have a responsibility to make value judgements about the futures they portray." (Veikko, Kanerva, and Kouri 2009, 9)

It should be noted this is not a universally accepted position within foresight studies, which typically distinguishes normative from exploratory studies (Coates and Glenn 2009). This introduces the possibility that characteristics were selected, or not selected, because they did, or did not, support a normative position the researcher wished to promote. This concern will be dealt with in more detail below.

The result of these deliberations was a simple schema for the identification of a species of ICT composed of the following elements:

1. Technology Name
2. History and definitions
3. Defining features (how it changes the way humans interact with the world)
4. Application areas and/or examples
5. Taxonomic relationship to other ICT's
6. Ethical and other concerns
7. References

(Stahl 2011b)

3.5.2. Step 2: Identification of Technologies

Technology identification proceeded by (human) textual analysis of 27 sources. The information was recorded in an online database. This database is no longer available, nor have its contents or schema been published, so it is not possible to interrogate this stage of the project as a data source nor validate it by attempting to reproduce the work. ETICA documents reference a number of Work Packages explaining the processes in more detail, but these are no longer available. The research design called for population of this database via analysis of governmental documents related to funding of ICT projects and of documents from "scientific/implementation" sources (Ikonen et al. 2010, 5). No precise definition is given for this phrase, nor is it apparent from the list of sources in what manner 'scientific' and 'implementation' are linked together, or what is meant by 'implementation.' The surveyed documents came from foresight studies organisations (such as the RAND corporation), technical standards bodies (such as the European Telecommunications Standards Institute), NGO's (such as Greenpeace), commercial enterprises (such as Microsoft), research and commercialisation organisations (such as the National ICT Research Centre of Australia), and EU organisations (such as the FP7 Funding program). No information is provided regarding on what basis these particular documents were selected, whether work was done to determine how typical they were, or whether other candidates were considered but discarded (and if so, on what basis).

Each candidate technology's record in the database contained a "general description" which "could contain" a number of fields (Heersmink, van den Hoven, and Timmermans 2010, 4). It is not stated in the ETICA documentation whether all fields were populated in all records, so we cannot determine whether all technology candidates contained the same data points.

The elements in the general description were:

1. Technical System (e.g. ICT, biotech, nanotech)
2. Field of Application (e.g. ageing, automotive industry, environment)
3. Target Audience (e.g. children, consumers, scientists)
4. Budget (numerical, if available)
5. Time Scale (numerical, if available)
6. Source

Additional fields could be attributed to each record. These were described as falling into four classes:

1. Social Impact (e.g. access, security, economic consequences)
2. Critical Issues (i.e. ethical issues, such as data protection, freedom, employment)
3. Capabilities (e.g. connectivity, interoperability, miniaturisation)
4. Constraints (e.g. complexity, reliability, safety).

Each class offered the researcher a limited range of choices and the option of a free-text entry.

This data structure was termed an "analytic grid" (Ikonen et al. 2010, 3) and represented in many ETICA documents with the following diagram:

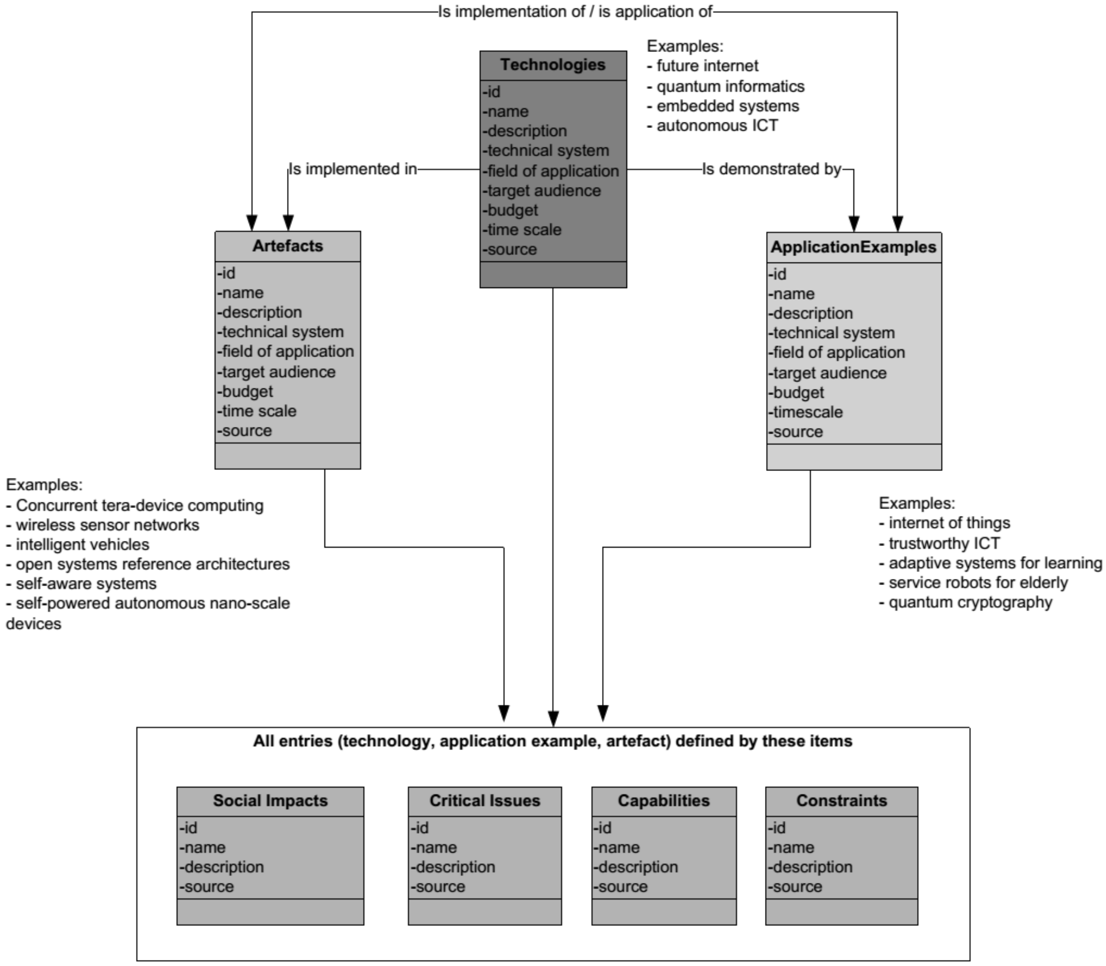

Figure 1 ETICA analytic grid (Ikonen et al. 2010, 3)

A number of questions arise as a result of the loss of the database and the lack of detail in the descriptions of it contained in the ETICA documentation. These will be dealt with below.

This stage of the analysis identified 107 different technologies. This data was cross-referenced against a list of technologies compiled from ten foresight studies research publications, eight from within the European Union and two from the USA. No information is provided as to the methodology by which these ten documents were selected. The resultant data was then condensed to a more limited list of higher-level technologies. Finally the technological identification data was standardised by the production of "meta-vignettes," described as "encyclopaedia-like" records containing "sufficient information" for ethical analysis (Heersmink, van den Hoven, and Timmermans 2010, 8).

The use of the term "meta" by ETICA indicated that these vignettes were to describe the technology in question at a high level of abstraction, rather than offer descriptions of sample artefacts, situations or other case-specific examples. The meta-vignettes contained a history of the technology, example applications drawn from the source literature, defining characteristics induced from these examples which focused on emerging features, related technologies and critical issues identified within the source literature. Of critical importance for understanding the methodology behind the creation of the meta-vignettes is that researchers had the option to supplement the database with their own sampling of "pertinent" literature (Ikonen et al. 2010, 8). What literature was sampled and the criteria by which it was selected is not documented.

3.5.3. Technologies Identified

The project identified the following eleven technology groups:

- Affective computing
- Ambient intelligence
- Artificial intelligence
- Bioelectronics
- Cloud computing
- Future internet
- Human/machine symbiosis
- Neuroelectronics
- Quantum computing
- Robotics
- Virtual/augmented reality

(Ikonen et al. 2010; Stahl 2011b)

ETICA found the developmental trends in these technologies exhibited some common characteristics, which may be seen as the features of trends within ICT development in general; convergence; increasing penetration of daily life, reusability and adaptability, reduction in size, increasing mediation of human perception and rising intelligence (Stahl 2011b). Note that many of these trends detected by ETICA are referenced above as some of our own research premises (see *Premises in ICT innovation, p.25*).

The following summaries of ETICA's technology groups are taken from *D.1.2 Emerging Technologies Report* (Ikonen et al. 2010).

3.5.3.1. Affective Computing

ETICA fused emulation and understanding of human emotion under the same term. In our view this is a problematic fusion of two distinct technologies, with distinct challenges, features and applications. While it is possible to develop systems which both interpret emotion and emulate it, the essential definition of technology provided was that it altered the way humans interact with the world. The ability of ICTs to understand human emotion produces a fundamentally different set of changes from the ability of ICTs to offer emotional output to humans. Fusing these two does serve to limit the complexity of the taxonomy, but makes it difficult to talk of affective computing as if there were a common set of features and concerns.

The primary features of this technology which form the foundations of ethical concern relate to the centrality of human emotion in communication and thought and the consequential risks of manipulation and misunderstanding. The critical issues identified were that ICT developers lack sufficient understanding of what emotion is, how it works, and its role in human society; and the encroachment on human autonomy of affective computing as a result.

3.5.3.2. Ambient Intelligence

Ambient intelligence is technology which embeds input, processing and response into the environment. ETICA identified the primary defining characteristic as the invisibility of the input devices, often associated with involuntary input (for example, a change in room temperature) and less-than obvious responses by the ambient system (for example, changing the air conditioning settings). Ambient intelligence was characterised by six features:

- Embeddedness: Ambient technology is embedded into the environment and may not be noticeable.
- Interconnection of ambient devices into complex distributed multi-function systems.
- Adaptive networking.

- Personalisation to the characteristics of each user.
- Anticipation of user needs or desires, with consequent response.
- Context Awareness[8].
- Innovative interaction paradigms.

3.5.3.3. Artificial Intelligence

It is unlikely Artificial Intelligence fits within ETICA's research scope of having a significant influence on society by no later than 2026. While reference is made to the distinction between "hard" and "soft" AI, the primary focus was on the emulation of human thought. Soft AI was only referenced in terms of expert systems. The features underpinning ethical concerns were competition with humans by artificial intelligences, the implications for law and human autonomy of automated decision processes, and the unpredictable consequences of independent machine learning.

3.5.3.4. Bioelectronics

Bioelectronics was defined as ICT systems which interact directly with the human (or other animal) bodies. Primary ethical dimensions relate to the use of ICT systems in health care but reference was also made to interest in the ultimate potential of bioelectronics to provide artificial human bodies. Here we can see that bioelectronics positions human physicality on a continuum with inorganic ICT systems and offers the potential for technological modification of the human form. It therefore directly confronts issues related to the essential and potential nature of the human and concerns regarding human autonomy.

3.5.3.5. Cloud Computing

It is notable that ETICA's definition of cloud computing is simpler and less technically nuanced than those of technical bodies, such as the US National Institute for Standards and Technology's *NIST Definition of Cloud Computing* (Mell and Grance 2011). Cloud computing may be seen as a re-working of the traditional client-server networking model (Voorsluys, Broberg, and Buyya 2011). Though ETICA's researchers did not use these terms, ETICA's characterisation distinguished cloud computing from traditional client-server network models by reference to the global nature of current cloud computing services, complexity of service provision, use of the Internet and an implied

[8] Incorporation of context as input and as processing variables.

strong scalability. In this sense, ETICA's characterisation of cloud computing can be described as primarily based on economic models and current (contingent) deployment practices, rather than the technology functions used when defining other technology groups. As a result, it is possible some of ETICA's ethical assessments regarding cloud computing are reflections of current economic circumstances or deployment practices as opposed to necessary characteristics of the technology itself. This concern is reinforced by the characteristics identified in the technology identification as of ethical concern. With the exception of privacy and data ownership issues, the features of concern are not so much ethical as engineering, such as quality of service and security. This issue is dealt with in more depth below (see *4.2 Key Dialectics in Cloud Services, p.74*).

3.5.3.6. Future Internet

ETICA used the concept of Future Internet as "a higher level concept that refers to the Internet in the future in general regardless of what it will be like" (Ikonen et al. 2010, 80). The functional aspects of Future Internet described in the *Emerging Technologies Report* focus on issues related to the Internet as a (complex) transport mechanism. ETICA's assessment focused on current trends relating to the connectivity of non-traditional devices to the Internet, the development of increasingly complex services and the underlying communications protocols. As such, the features of Future Internet which give rise to ethical concern overlap those of ambient intelligence and cloud computing, most especially privacy, intellectual property and transparency.

3.5.3.7. Human-Machine Symbiosis

ETICA characterised human-machine symbiosis as the pairing of natural human capabilities with ICT. This definition covers an extremely large range of artefacts and ICT systems, including virtual and augmented reality, expert systems and direct neural interfaces. Thus, as with Future Internet, there is considerable overlap with other technologies, especially bioelectronics and artificial intelligence.

3.5.3.8. Neuroelectronics

> "Neuroelectronics, sometimes referred to as neurotechnology, is the discipline that deals with the interface between the human nervous system and electronic devices. It is a highly complex and interdisciplinary field with contributions from computer science, cognitive science, neurosurgery and biomedical engineering. Neuroelectronics has roughly

> three related branches: (1) neuroimaging, (2) brain-computer interfaces, and (3) electrical neural stimulation… The only defining feature these three branches have in common is that they all interface electrical devices with the brain, either to extract information *from* the brain or to send electrical signals *to* the brain." (Ikonen et al. 2010, 98–99)

ETICA focused on neuroelectronics mainly as a mechanism for using the neuronal activity of the human brain as input, either for control systems or for imaging. Significantly less discussion was devoted to output directly to the brain. However, discussion of critical issues (which serves as the basis for ethical concerns) focused on the use of neuroelectronics as a means of processing thought, on the highly problematic assumption that "neural signals may indicate - even represent - thoughts" (Ikonen et al. 2010, 100). This potentially limits the ethical considerations within a fairly narrow and highly contentious materialist view of humanity and simplifies to an extreme degree the difficulties associated with developing an account regarding how neuronal activity gives rise to the phenomenological experience of thought (Halliday 1998), the attendant metaphysical and epistemological difficulties and their impact on ethical considerations.

3.5.3.9. Quantum Computing

No definition of what constitutes quantum computing is provided by ETICA. However, the discussions of quantum computing in the ETICA documents imply a broad definition of "computational machines that use … quantum effects" (Ikonen et al. 2010, 103). Quantum computing clearly falls outside the temporal scope of the ETICA project, and is indicated as such within the ETICA reports. ETICA did not consider it likely, or even plausible, that quantum computing would have a significant impact on society by 2026. However, no explanation is provided as to why it was nevertheless included in the ETICA study. Its presence in the technology taxonomy enhances the value of that taxonomy as a general description of current areas of ICT activity, but offers nothing towards the ethical concerns which form the focus of ETICA's activities. No critical issues were identified which could give rise to ethical issues of concern today.

3.5.3.10. Robotics

ETICA defined robots as "machines with motor function that are able to perceive their environment and operate autonomously" (Ikonen et al. 2010, 114). It is worth noting this includes self-driving cars, which were not anticipated by ETICA.

ETICA's characterisation focused on new developments which increase the mobility and intelligence of robotic devices, permitting their deployment into wider areas of society, such as the home and healthcare. Application discussions focused on robots built for specific and narrowly defined roles within these contexts, rather than more speculative concerns about theoretical general-purpose, fully mobile devices controlled by strong artificial intelligence. However, discussions of the critical issues which could found ethical concerns reversed this emphasis to focus on the latter rather than the former, such as the danger of the human race being taken over by robots.

3.5.3.11. Virtual/Augmented Reality

ETICA's union of virtual reality with augmented reality into a common technology was based on the common feature shared; the imposition of digital output into human sensory input. Virtual reality occurs where that digital output substantially or completely replaces sensory data from outside the ICT system. Where it does not it is termed 'augmented reality.' ETICA's concerns were founded on the fact that such systems mediate or replace interaction with the physical environment.

3.6. Assessment of ETICA's Technology Findings

3.6.1. Assessment Methodology

There are many methodologies for assessing qualitative research, some competing and some complementary. Many have supported this state of affairs with a variety of arguments in favour of such a pluralist approach (Stige, Malterud, and Midtgarden 2009). The case for the pluralist approach is further strengthened if one accepts that evaluation of foresight research, in particular, cannot be independent of the context and purpose for which it was undertaken (Georghiou and Keenan 2006). One could further argue that the pluralism of research assessment methodologies is appropriate to the pluralism of foresight research methodologies. For example, *Futures Research Methods* identifies thirty-nine distinct methodologies within foresight research (Glenn 2009b). A variety of assessment approaches facilitates the selection of the assessment methodology best suited to a specific research methodology. However, the selection of a particular assessment methodology also brings with it concerns as to the basis for that particular choice. The selection of an assessment methodology is not, in and of itself, a neutral or automatic choice. Dominant paradigms within particular fields, such as that of qualitative scientific experimentation, may bias the selection of assessment criteria (Northcote 2012; Hannes 2011). In addition, the individual assessor's own understanding of critical assessment criteria, such as trustworthiness, reliability and validity, may vary according to their understanding of the research material itself, or their background philosophical understanding of these concepts or even of the assessment process itself (Krefting 1991). Furthermore, the method by which one selects an assessment methodology can, itself, be subject to the same forms of scrutiny. Thus is becomes possible to first assess the of the choice of assessment, and then to assess the assessment of the choice of assessment, and thus fall into an infinite regress of assessment of assessments. Such a process is clearly to be avoided by mere mortals with finite lifespans. Nevertheless, it is important to have a clearly defined methodology by which to assess foresight research projects in order to ensure assessment covers all relevant aspects of the research. Furthermore, an explicitly documented assessment methodology exposes assumptions regarding what are the important aspects for evaluation, why and by what methods, so that the assessment of the research may itself be evaluated (Kothari 2004; Lakatos 1970).

Our methodology for assessment of the technology identification phase of ETICA's research will seek to address these concerns by selecting an assessment methodology which is of general application and which is used by the widest possible audience. Such an assessment methodology has several merits. In that it is general and intended to assess a wide variety of research projects, it must be limited in the degree to which it relies upon, or reflects, specific methodological approaches or paradigms, either in the assessment or the research itself. Secondly, a methodology which is used by a wide range of assessors must be limited in the degree to which it reflects particular cultural, disciplinary or philosophical premises and practices. Finally, we accept the argument that an assessment needs to give some space to the peculiarities, or even flaws, of a specific research project's methodology in order to avoid obscuring the value of its findings (Edwards et al. 2000). This last point argues in favour of an assessment methodology with a "loose fit" to the research methodology in questions. It is also likely that such a neutral approach, in not being too closely tuned to a specific research methodology, will better accommodate research projects which utilise multiple methodologies. We shall therefore use a framework adapted from the Critical Appraisal Skills Program (CASP) developed in Oxford in 1993 by Dr. Amanda Burls (Dixon-Woods et al. 2007).

The CASP methodology has been widely accepted in public health and related circles for the methodological assessment of research and is now widely used in many countries (Singh 2013). A review of journal publications reveals many articles explaining how to apply it to specific fields of medical research, which we have taken as evidence that the CASP methodology provides satisfactory working methodology for many professionals, including those upon whose decisions lives depend. In addition, the use of the CASP methodology has spread and continues to spread, which we take as further evidence of its utility. While the methodology originated, and is most popular, within medical research, there is nothing within the methodology itself which is restricted to any specific research discipline.

3.6.2. Assessment Checklist

CASP research assessment methodologies consist of series of questions (typically referred to as 'checklists') designed to ensure appraisal covers the full range of assessable variables by which to determine the quality, trustworthiness, methodology, context, implications and importance of any particular research project. The CASP methodology can be therefore profitably applied to assessments of any form of methodological research, qualitative or quantitative. There are multiple versions of these checklists available, none with canonical status. Some variations reflect increasing sophistication of approach over time, in that older checklists are often shorter than more recent ones. In addition, some have been tuned for specific fields of medical research and contain questions meaningless outside their medical speciality. Our protocol for selection was that checklists should contain questions which were applicable outside medical research; that they should come from reputable organisations; that they should be intended for an audience actively engaged in assessing research and which was compelled to use these checklists, thus ensuring they *were* actually using them. Our rationale for this last requirement was that we can assume the checklists in question are capable of being used to fulfil their purpose when they *are* being used for said purpose *and* enough time has elapsed for them to have been modified if they were unfit for purpose. Furthermore, we have drawn checklists from two such sources and combined them. The rationale for doing so is to increase the chance that relevant questions will be included in our final version by drawing from more than one data source. In addition, doing so provides the opportunity to fill gaps in the areas covered should a single checklist need to have questions removed because they are too specific to the medical field. To this end checklists from ten different sources were examined. However, we have chosen only two sources from which to draw checklists; the UK National Institute for Health and Care Excellence (NICE), which lays down operative regulations for the entire UK health service, and the UK Civil Service. It was found that other sources either copied from these sources exactly, simplified their checklists so that they contained fewer questions, or modified the questions to make them specific to a particular branch of research. The NICE checklists were drawn from the UK's National Institute for Health and Care Excellence (NICE) *Guidelines Manual Appendix B: Methodology Checklist: Systematic Reviews and Meta-Analyses* (NICE 2012) and *Guidelines Manual Appendix H: Methodology Checklist: Qualitative Studies* (NICE 2007). Our rationale for selecting these two checklists is that it is appropriate to treat the bibliometric and textual analysis as a form of systematic review, while the surveys and the focus groups all constitute examples of qualitative research. The Civil Service guidelines, *10 Questions to Help You Make Sense of Qualitative Research* (Critical Appraisal Skills Programme (CASP) 2006) are widely distributed within many UK government departmental websites in PDF format.

We have combined the checklists and removed several questions applicable only to medical research on human subjects. This yields the following questions for a review of the Phase One of the ETICA project, the identification of ICT's of concern:

1. Was there a clear statement of the research aims?
2. Were qualitative and quantitative methodologies used as appropriate?
3. Are the roles of researchers clearly described?
4. Was the research design appropriate to address the aims of the research?
5. Was the sampling strategy used appropriate to the research question(s)?
6. Are methods of data collection adequate to answer the research question(s)?
    a. Was an adequate description of the application of the methodology provided?
    b. Was the literature search sufficiently rigorous to identify all the relevant studies?
    c. Were the search and collection methods used appropriate to the research question(s)?
7. Was the data analysis sufficiently rigorous?
8. Are the findings relevant to the research question(s)?
9. Are the findings internally coherent, credible (valid)?

10. Is there a clear statement of findings?
11. Is there adequate discussion of the study limitations?
12. How valuable is the research?

We have modified the wording of some questions to cater for quantitative methodologies and the bibliometric analysis. Additional supplementary questions have been added to Question 6 (data collection) regarding the review of existing journals and other publications. These were adapted from the NICE *Guidelines Manual Appendix B: Methodology Checklist: Systematic Reviews and Meta-Analyses* (NICE 2012).

3.6.3. Assessment Questions

1. Was there a clear statement of the aims of the research?

A formal and explicit statement of a specific research question or set of such questions does not exist in any ETICA publication. Reasonably clear statements of the project's aims can be found in most deliverables, but they tend to be nuanced to suit the topic of each particular deliverable. We have selected a sample set of five variations in the statement of research aims from different ETICA publications to serve as data points for our analysis:

Sample A: *Deliverable D.1.2 Emerging Technologies Report*:

> "The overall aim of the ETICA project is to identify emerging ICTs and the ethical issues these are likely to raise with a view to providing policy recommendations to the EU as well as other policy makers." (Ikonen et al. 2010, 2)

Sample B: *Deliverable D.4.2 Governance Recommendations*:

> "The ETICA project aims to analyse emerging technologies in order to evaluate ethical issues that arise from them, rank them, and assess current governance approaches to these issues in order to determine the nature of current ethical governance. It aims further to formulate new ethical governance recommendations based upon interpretations of these analyses in order that ethical issues can be addressed before or as they arise." (Rainey and Goujon 2011, 7)

Sample C: ***D.1.1 Heuristics and Methodologies Report*:**

> "The Ethical Issues of Emerging ICT Applications (ETICA) project will move beyond the state of the art in ICT ethics by offering a comprehensive overview of ethical and social issues in ICT that are likely to develop in the medium term future. ETICA will do this by providing a list of emerging technologies and their applications as well as develop a matrix of technological applications within which ICT ethical issues can be easily identified. The potential ethical issues will be subsequently ranked and graded by evaluating them according to top priority issues." (Veikko, Kanerva, and Kouri 2009, ii)

Sample D: ***Final Report*:**[9]

> "The ETICA project identified ethical issues of emerging ICT applications. … It also described ethical issues related to these technologies. Furthermore, the project described the methodology used to arrive at these findings. On the basis of the identified technologies and their ethical aspects, the ETICA project evaluated and ranked them. It investigated current and possible ways of implementing governance that are conducive to addressing the ethics of emerging ICTs." (Stahl 2011b, 1)

Sample E: ***D.1.3. Emerging Technologies Workshop Report***

> "The workshop brought out clearly the vast challenge of ETICA project: its ultimate goal is to construct feasible and usable, concrete tools or framework for anyone to explore ethical issues when developing and designing future ICTs." (Veikko, Panu Kouri, and Ikonen 2011, 5)

Sample A can be summarised as a) identify technologies b) identify ethical issues c) make recommendations. Sample B contains an additional aim of assessing "current ethical governance." This aim is also found in Sample D, but not samples A, C or E. It does, however, find expression in formal ETICA output, *D.4.1 Governance Approaches* (Goujon and Flick 2011), which addresses the "characteristics and limitations"

[9] The *Final Report* does not contain an explicit account of ETICA's aims. Instead it contains a summary of what ETICA did. We have included it in our sample because, as the final report, it constitutes the most authoritive account of the ETICA project.

(Goujon and Flick 2011, ii) of current ethical governance. Sample C introduces "social issues" as an additional area of concern not listed in the other samples. Sample D adds the publication of methodologies. Sample E's term 'framework' summarises the taxonomic developments referenced in Samples B, C and D, and brings further understanding to Sample A's term 'identify.' We can therefore understand Sample A's phrase "identify emerging ICT's" as creating and populating an organising structure ("framework") and developing methods for dealing with it ("tools"). This reference to tools is reflected in Sample D's reference to publication of methodologies, which would support an aim of developing tools. Furthermore, the majority of the content of ETICA's publications are concerned with methodology, not findings. Thus, methodological development constitutes a significant, possibly the main, research output. In that this is reported as the aim project participants had in mind when participating in a project workshop, it is reasonable to assume this aim was operative throughout the project. However, it is, at best, implied, but not explicit, in most samples.

In addition to the imposition of additional aims in different statements, some of the terms used are ambiguous. It is not clear what constitutes an "evaluation" of ethical issues. This could, for example, be nothing more than a simple ordered list of ethical topics arranged according to a simple set of criteria. Conversely it could be an in-depth exploration of the philosophical thought relevant to each issue. Sample C's reference to "social issues" is similarly problematic in that it is far from clear what is meant by the term, or by what discipline or methodology they will be examined; political, philosophical, or sociological or some other.

It is therefore our conclusion the ETICA project did not provide a clear (and therefore unambiguous) statement of the project's research aims. While there is considerable overlap in various descriptions, there are subtle nuances of difference which, when guiding the formation of a full report, can easily become magnified into large variances of content, premises, arguments or other elements with a determinant effect on analysis and conclusions. It is further our conclusion that a single clear, short and standardised statement - formally stated - of the research question(s) would have benefited the project in two ways. First, such a statement would increase ease of understanding for those reading the reports and thus facilitate effective communication and reception of project findings. Secondly, in terms of project operational efficiency, such a statement could have served as a common reference point by which to orient all

ETICA researchers. It is likely the process of formulating such a statement would have necessitated discussion amongst the ETICA participants in order to achieve consensus. This would have helped to ensure a common understanding of the projects aims to a detailed degree amongst all project participants. This could have provided additional potential benefits by promoting more uniform communication of the project by researchers to subjects during the various focus groups and surveys.

2. Were qualitative and quantitative methodologies used appropriately?

ETICA devoted considerable effort towards methodological exactitude. Checks were put in place at each step of data gathering or analysis to verify methodological compliance. Selection of appropriate methodological procedures were the topic of a specific deliverable; *D.1.1 Heuristics and Methodologies Report* (Veikko, Kanerva, and Kouri 2009). This document demonstrated a concern for bias within the raw data itself, constraint of data sources via reference to project aims, justification for the data sources selected being consistent with intended methods of analysis and with project aims, and a number of further verification steps, such as surveys and focus groups with which to cross-reference ETICA researcher findings. These concerns are reiterated in depth in other deliverables, most notably *D.3.2 Evaluation Report* (Rader et al. 2010), which provides detailed justification for the methodologies chosen for each phase of the research.

The other aspect of appropriate choice of quantitative versus qualitative methodologies arises regarding the data to be processed. ETICA's raw data sources were textual. To analyse this data quantitatively would involve bibliometric analysis. Qualitative analysis would require human reading of the text and reflection upon it, the traditional form of textual analysis and the only one available until recent years. It is notable that ETICA used both forms of analysis in a complementary fashion, such that they supported each other.

It is our conclusion that the combination of methodologies, the reasoning in determining their usage, and the documentation of methodological procedures, can be regarded as a paradigm of foresight research. The combination of methodologies with the quantity of data analysed has been cited elsewhere as a model of foresight research (Brey 2012; Stahl, Mittelstat, and Fairweather 2013; Pellé and Reber 2013), which constitutes further evidence of the paradigmatic status of this project's methodologies.

3. Are the roles of researchers clearly described?

While it is clear from ETICA documentation which functions ETICA researchers undertook, it is less clear what approaches and operational methodologies each used in performance of their individual tasks. Of particular concern are occasional references to normative imperatives, suggesting the possibility that the normative values of researchers could have affected data selection, gathering, processing or the development of recommendations. For example, the *Heuristics and Methodologies Report* states:

> "The practice of futures research has always a normative aspect to it because each scenario, forecast and image of the future have a potential and usually also an effort to affect the future. Therefore futurists have a responsibility to make value judgements about the futures they portray." (Veikko, Kanerva, and Kouri 2009, 9)

This is reiterated in *D.4.1 Governance Approaches*:

> "When we talk about the future, it is impossible to know the context that the future might be framed by. This means that we cannot really analyse it except through some sort of normative approach ... we have approached the problem [by] investigating the possible approaches that are available, from a normative perspective, to address issues found in future conditions." (Goujon and Flick 2011, 1)

A normative approach would have the greatest influence on the more explicitly ethical portions of the ETICA project; the ethical considerations and the governance analysis. However, there remains the possibility that normative stances may have influenced the technology analysis itself. *D.4.1 Governance Approaches* states that the foundation for the normative approach is the work of Lenoble and Maesschalck, advocates for specific and innovative changes in law and governance, especially with regard to public-private interactions and issues of legislative legitimacy, and who work from a pragmatic theory of law (Lightcap 2011). It is beyond the scope of this chapter to critique their work. However, it is important to note our concern that Lenoble and Maesschalck's work presupposes an essentially static structural and institutionalised power arrangement and may have insufficient regard for the dynamics of power within social discourse (Lightcap 2011). In that this could promote a skewed selection of technological trends, this could lead to bias in selection of the data points for analysis leading towards technology identification. Critical technologies can emerge from sources or methodologies which are not dominant within the power structures of their field at the time of their development (Day and Schoemaker 2004; Curran, Fenton, and Freedman 2012; Limonard and de Koning 2005; Assink 2006; G. M. Schmidt and Druehl 2008). Their effect is often to upset existing economic or power structures. It is for this reason they are termed "disruptive" technologies. It can be argued, therefore, that one needs to take into consideration power structures underlying those who provide accounts of technology trends (Curran, Fenton, and Freedman 2012; Hillis, Jarrett, and Petit 2013) and correct for sampling bias by drawing data from non-institutionalised sources. For example, one could seek to explore technology developments within de-institutionalised funding, such as crowd funding (Belleflamme, Lambert, and Schwienbacher 2014). Alternatively, one could survey the product aims of start-ups populating the multitude of digital incubation hubs found in many cities (Tötterman and Sten 2005).

We must therefore conclude that the influence (if any) of the researcher's norms on the project are indeterminable. However, this may be said of most research projects. All aspects of a research project must inevitably be influenced by the premises and normative values of the researchers. Indeed, all research depends upon the participants being influenced by key normative values such as the desirability of accuracy. Much of the emphasis in any methodological design is focused on minimising such background influences. However, it is not practically possible to do this to an absolute degree, so some uncertainty must always remain. The concern for normative bias in the identification of emergent ICT does not therefore, in and of itself, invalidate the resultant findings. It may, however, suggest an area which could profit from further development. It may be possible to refine the technology projections by greater reference to power dynamics in technology innovation, or by contextualisation within a different theory of law. Alternatively, it is possible that such approaches would yield conflicting results. These possibilities do not constitute flaws in the ETICA project, for all research projects contain areas of work suitable for further development, and thus are able to serve as foundations for further research.

4. *Was the research design appropriate to address the aims of the research?*

Assessing whether the research design was appropriate to the research aims is complicated by the lack of a single clear and explicit statement of those aims. Nevertheless, there is sufficient overlap between the different versions, and they possess sufficient accord with what the ETICA project actually did, to make such an assessment.

The aims of the ETICA project were to enhance the capability of the EU's private and public sectors to anticipate and respond to ethical developments stemming from ICT innovation. This was accomplished by providing three research outputs. Firstly, the project organised understanding of the field of emerging ICT's. This was accomplished by developing conceptual frameworks which organised the field of enquiry and provided tools of description and analysis. These were designed with an emphasis on features which would support ethical analysis. This aim was successful in that they did serve as the basis for development of ethical issues. In parallel with the development of data and data structures, methodologies were developed by which data processing occurred. Two complementary taxonomies were constructed; a taxonomy of emerging ICT's, from which a second taxonomy (of ethical concerns) was developed. The third and final output was a review of the existing state of ethical governance and consequent recommendations regarding the integration of the data and methodological output into the ethical governance processes of EU public and private sectors.

It is clear that the research design supported these aims because the project produced several thousand pages of relevant output spanning fourteen reports. However, some variations of emphasis between aim and design may be problematic. It was stated that ICT analysis "needs to be associated with a vision that embodies specific views of humans and their role in the world" (Ikonen et al. 2010, 4; Stahl 2011c, 63). In addition to its relevance at this stage of the research, this is important during ethical consideration because said vision largely determines the intent of the actor, which impacts on moral culpability and other ethical dimensions under many ethical systems. However, it is not clear that this consideration was actually a factor during analysis because it is not discussed with reference to the technologies identified in any of the deliverables, nor is it formally represented in the analytic grid or technology accounts.

An additional area of concern lies in the view taken of what constitutes technology. Here is the greatest area of methodological weakness. What constitutes technology at this level of analysis is far from indisputable (Reyden 2011). ETICA's stated definition of technology was as "high-level socio-technical systems that incorporate a view of humans and have the potential to change the way humans interact with the world" (Stahl 2011b, 11). However, an examination of the analytic grid reveals no elements for the recording of social, economic or similar characteristics. The technologies are in fact, described in terms of their operational purpose and their instantiation in artefacts. Social factors are mentioned in some publications, such as *IT for a Better Future* (Stahl 2011a), but they are discussed only in relation to ICT's in general and not as characteristics of individual ICT's. Hence two areas of departure occur between aim and methodology. First, as a disputable definition of technology it would have been appropriate to devote more emphasis to explaining and defending this position. However, this is a minor point. Secondly, and of greater concern for this aspect of analysis, is the lack of socially-related characterisation of technologies. It is frequently the case that social aspects of a technology system, such as patterns of user reception, influence not just how a technology influences society, but also on how that technology itself develops (Limonard and de Koning 2005; Hillis, Jarrett, and Petit 2013) while economic power structures exert a similar influence (Bogliacino and Pianta 2013; Fontana and Nesta 2009). In both cases we see not only the impact of factors contextualising the technological artefacts, but also feedback loops illustrating the systemic quality of ICT innovation. Indeed, in the previous discussion regarding ETICA's characterisation of ambient intelligence (see *p. 45*) we noted an apparent conflation of contingent economic structures with necessary technical characteristics.

We must conclude, therefore, that while the initial design defined technology as a socio-technical system, this definition was not fully reflected in the design of data structures, which lacked elements in which to record both social and systemic data. We consider the absence of social data to be of greater concern for ETICA's findings than the absence of consideration of ICT's as systems, because of the immediate and direct relevance of social data for ethical and governance issues.

5. *Was the sampling strategy used appropriate to the research question(s)?*

In some ways ETICA's approach to sampling documents shares the same logic as the scenario development seen in the meta-vignettes. In using as the core data multiple papers from multiple authors in multiple publications, ETICA was not wedded to a single scenario regarding the future. In fact, it was "expectationally neutral" (so to speak) in that assessments of ethical concerns within the source papers took no account of the futures anticipated by the authors of these papers. This is an accepted approach within foresight studies regarding specific visions of the future (Glenn 2009c). However, no attempt was made to consider alternative sources outside mainstream institutional structures. This issue is relevant because it speaks to the representativeness of the sample. The delimitation of data sources was justified on the grounds it would:

> "Give a good view of what is intended and envisaged by organisations that are in a position to enforce their view of the world." (Stahl et al. 2010, 30)

Elsewhere the justification for source provided is that:

> "It was felt that governmental sources and research sources would provide a more reliable picture of emerging ICTs. Such sources are related to research and development activities and therefore less likely to be speculative than other future-oriented research, which can be motivated by aims to raise attention or create markets." (Ikonen et al. 2010, 17–18)

However, ICT innovation has a clear pattern of frequent and significant user-driven invention, in which people use a technology in a manner not anticipated by its inventors (G. M. Schmidt and Druehl 2008). This suggest that academic, corporate and other large institution's understanding of ICT trends is often unable to anticipate innovative impetus from non-institutional actors (Day and Schoemaker 2004). It is therefore appropriate to consider the wider social setting when selecting sample texts and, if selecting a sample from one discipline, to consider the variation between that sample and the general population, to what degree it is representative, and how this is demonstrated, or if not demonstrated, how that is accounted for, during data processing and analysis. While it is sufficient to maintain result validity by qualifying predictions with the fact they are from a limited range of possible sources, the report produces policy and governance recommendations to legislative bodies and so should, if it does serve the purpose for which it was intended, produce real effects within society. It is therefore important that the original sample set represent the full range of relevant social perspectives.

We have indicated above concerns regarding the limitation of the sample set to institutional sources and the potential value of examining other sources of ICT innovation (see *Are the roles of researchers clearly described?* p. 57). The value of drawing from a wider range of sources is that it draws on wider society's technological imagination. As the power of technological innovation grows so the proportion of society which can harness it and contribute to ICT innovation grows, and so the range of causative agencies of ICT development expands beyond the traditional institutional structures and processes. The popular social imagination is a significant and meaningful source of data because it represents the imaginative foresight of the larger group of innovation initiators existing outside mainstream institutions. Such foresight considerations are most easily seen in science fiction movies and books, which offer analysis of near future ICT's and the ethical implications arising therefrom; such as the TV series *Continuum* (Barry 2012), the movies *Minority Report* (Spielberg 2002), *The Bicentennial Man* (Columbus 2000) and *Blade Runner* (Scott 1982), and works of literature such as *Brave New World* (Huxley 2006), *1984* (Orwell 1983) and the works of many other science fiction writers. Indeed, criticism was directed at ETICA in this respect before the project had finished (Christensen 2009). It is impossible to know if analysis of fictional sources would have yielded additional insights, contradicted the present findings, or merely confirmed them. It remains, however, clear that a potentially valuable source of data was untapped. Only a comparable exercise on such data would reveal to what degree failing to use it weakened ETICA's output, if at all.

6. *Are methods of data collection adequate to answer the research question?*

a. *Was an adequate description of the application of the methodology provided?*

The algorithms and data structures used by VOSViewer are not documented within ETICA, though much of this material is available in external publications (Van Eck and Waltman 2009, 2007; Heersmink et al. 2011). However, the underlying data structures giving rise to the cloud map images has not been published and is not available. This makes it impossible to reproduce or validate this step in the research. Particularly problematic is the lack of clear information regarding the dimensions of the bibliometric analysis. External publications regarding VOSViewer provide general information regarding its algorithms (Van Eck and Waltman 2009, 2007) and its implementation within the ETICA project (Heersmink et al. 2011), but these are not sufficient to enable reproduction of the VOSViewer phase of the research.

The loss of the online database for technology determination is also unfortunate. The system contained a wiki which "reflected the dynamic of the decision process with regards to the technologies to be included in further analysis and was linked to the individual documents that contained the meta-vignettes" (Heersmink, van den Hoven, and Timmermans 2010, 21). Access to this wiki would have enabled examination of the reasoning behind the output.

b. *Was the literature search sufficiently rigorous to identify all the relevant studies?*

ETICA own evaluation considered the sample set limited in terms of being restricted to English-language documents of Western origin. These limitations are indisputable, though their impact is less clear. The source countries from which the samples were drawn are responsible for a considerable degree of the innovation in the ICT's identified, if not the majority. As such, other sources are likely to reflect innovative agents of less importance. It is the case that this weighting of influence in favour of Western English-language nations, such as the USA, is at least partly a feature of current contingent global economic and power structures. As such this dominance may change within the temporal scope of the project's projections. However, because the ICT data structures characterising the technologies lacked the capacity to hold such data, it is doubtful if considering changes in global innovative influence over the next ten to fifteen years would have been possible within the ETICA research design.

c. *Were the search and collection methods used appropriate to the research question?*

It is implied in *D.1.2 Emerging Technologies Report* (Ikonen et al. 2010) that the technology database was organised around the technology groups, rather than other features such as functionality, scale, or relationship with other technologies. This implies that determination of technology groups was made prior to analysis of distinguishing features and that determination of technologies was not induced from analysis of data. If this is the case, then technologies were identified by an unknown method, possibly varying from researcher to researcher, and then defining characteristics were identified which conformed to the pre-determined technological classification.

No definition is provided to explain what constitutes each of the items in the database description. In particular, no information is provided as to whether textual data, such as Field of Application or Target Audience, was constrained in any fashion or whether contributors were permitted to write whatever they wanted. This raises, but does not answer, questions as to the precision and consistency of the data points from which the technology identification was based.

Select lists at end of the data entry process may have been a methodological weak-point. These offered the researcher a limited range of choices and the option of a free-text entry. The list of what choices were offered and what they meant is not documented in the ETICA publications. It is not documented whether the researchers had guidance as to the meaning of the terms they selected amongst, nor whether any processes were in place to assess or ensure consistency of understanding amongst the researchers. However, each entry was peer-reviewed by other researchers within the project. Furthermore, since these were not the sole sources of the final technology groups, inconsistencies in understanding at this stage are unlikely to have had any significant effect.

7. *Was the data analysis sufficiently rigorous?*

VOSViewer only used a limited range of bibliometric techniques, primarily word frequency and proximity analysis, without synonym inclusion (Heersmink et al. 2011). This is but one form of bibliometric analysis and operates only at the word level. Of particular significance for an effort to gain a conceptual understanding of published discourse, VOSViewer does not take into consideration the semantic relationship between pairs of terms. Other dimensions of bibliometric analysis which could have been of value include some form of discourse level analysis, such as discourse structure analysis, which builds upon analysis of sentences or phrases to construct hierarchies of discourse segments (Ide 2007). In that a taxonomical hierarchy was constructed, co-referent analysis, which analyses the referencing structures of key terms in a document, could have provided bibliometric data in a hierarchical ordering which could have been mapped onto the ICT taxonomy.

It is clear bibliometric analysis could have offered more to the project. However, bibliometric analysis was not used to produce the final results of this phase of the research, but merely to guide the further work of human researchers. As such it is legitimate to argue that the modes of bibliometric analysis used were appropriately determined by the use to which bibliometrics results were to be put. From this perspective, other bibliometric research methods could only have yielded data which the project was not designed to analyse. Incorporating this data would have required considerable resources and thus substantially altered the research processes. There is no indication such a change would have yielded better results and so has nothing to recommend it. It does, however, illuminate an opportunity for a research utilising deeper and wider bibliometric analysis in this area.

The use of foresight studies material to cross-reference against the original technology database permits the technology database to act as a constraint on the material drawn from foresight studies in terms of timescales and practical application (ie: probability of emergence). As discussed above, the original selection of source material was based on the aim of understanding predictions and expectations of those who materially affect what technology emerges. The aim is to restrict consideration to ICT's more likely to develop. By contrast, foresight studies need not involve those who are involved in developing new technology, nor is it necessarily restricted to the possible, but may discuss technologies which may never come about or which may not be practical. As a research design methodology, it therefore becomes possible for ETICA to limit the foresight studies research used as source material in terms of timeframe and probability through use of the initial technology descriptions as topic filters. However, this is only possible if the following conditions are met;

- The original technology descriptions had themselves been strictly limited in accordance with these principles;
- The data structures encoded these limitations in a manner which permitted their reapplication during foresight studies analysis;
- Said data was incorporated as constraints when sampling the foresight studies sources.

The ETICA research methodology regarding these issues is not documented. However, we do find "speculative" technologies within the ETICA considerations in the form of quantum computing. This suggests some technologies were evaluated without reference to their probability of appearance within the twenty year timescale to which ETICA was limited. However, it is acknowledged that the emergence of this technology is outside the temporal scope of the project within the ETICA documents themselves (Ikonen et al. 2010). Many of the aspects of artificial intelligence discussed are similarly outside the project's timescale. This suggests the foresight studies review was not constrained by the findings in the initial technology descriptions. This in turn suggests the source material used for the final technology identifications was not selected according to the criteria listed by ETICA.

It is our assessment that this reduces the utility of the list of technologies as a complete definitive list of emergent technologies of concern upon which further foresight research can be based. Any future research which is based upon or uses the ETICA technology taxonomy needs to establish a timescale which matches that of the actual ETICA technologies, or refine the ETICA list to reduce the range of timescales for emergence. This does not, however, reduce the reliability of the technology descriptions or meta-vignettes themselves because the analysis conducted by the ETICA researchers introduced appropriate limitations of scope by confining the defining characteristics to those which within a ten to fifteen year scope. Since it was these defining characteristics which formed the data from which the ethical concerns were developed, aspects of a technology which were outside the scope of the project but present in the technology descriptions were filtered out of the research process at this point.

However, this introduces a concern that there may be a variance between the project's designed role for the bibliometric data and the actual use to which it was put. As indicated above (*see Step 1: Development of data schemas*, *p. 39*), it was considered that research bias may unduly influence meta-vignette creation and that the bibliometric results would function as constraints against this (Stahl et al. 2010; Brey 2012). However, we have demonstrated above that researchers were not limited by pre-existing data, but departed from it in order to correct for elements outside the project scope. We must therefore conclude that the bibliometric results did not function as a corrective for researcher bias during meta-vignette construction as the research design called for. In addition, as indicated above (s*ee Step 2: Identification of Technologies, p.41*) researchers could populate the database with their own sampling of "pertinent" (Ikonen et al. 2010, 8) literature. This adds an additional source of data to vignette creation, and an additional possible source of researcher bias, which was not subject to constraint by the bibliometric results. This leaves the project without any identified process for correcting researcher bias during meta-vignette creation. Determining to what degree meta-vignette construction was skewed by researcher bias is beyond the scope of this assessment. However, it is clear that the research design deemed the possibility sufficient to require a step in the research process to correct for it. It is also clear that this step was not implemented. As a result we must conclude that, with respect to meta-vignette construction, the procedures followed were not consistent with the research design.

8. Are the findings relevant to the research question(s)?

In that the aim of this phase was the development of a set of emerging ICT's, it is clear that ETICA produced a relevant list. However, while technology is defined as a socio-technical system, it is apparent from discussions above that the social aspects were not given equal weight to the technical within the data structures. It is also clear from discussions above that only a limited set of those involved in ICT innovation were sampled. It is therefore apparent that, while the findings were relevant, the design of the ETICA project called for additional relevant information which was not included. It is less clear, however, that such omissions damaged the findings to any significant degree. Determining the importance of such omissions is beyond the scope of this chapter and would, most likely, constitute a substantial research effort of its own.

9. Are the findings internally coherent, credible (valid)?

The taxonomy of technologies and associated descriptions are, for the most part, consistent. However, there exist occasional points of inconsistency in terms of compliance with the research design. These are rare and of limited significance. For example, quantum computing is identified as being beyond the temporal scope of the project, yet included. Here the application examples and explorations of social impact are far less detailed than those of the other technologies. Thus quantum computing is inconsistent with the other technologies in terms of both timescales and depth of analysis. A similar argument can be made regarding artificial intelligence. However, in that a limited number of technologies give rise to a very large range of ethical concerns at the next step in the project, and that multiple technologies can intersect at individual ethical concerns, it is unlikely that the inconsistencies detailed above give rise to significant degradations of reliability in the project's findings.

10. Is there a clear statement of findings?

This question is most important with regard to the project's final reports, which address ethical and governance issues, not the results of the technology identification phase. However, as we indicate below, the findings of the technology phase are of value in and of themselves. It is therefore appropriate to consider the accessibility of the results of *this* phase for future researchers. Here we regard accessibility of results to be largely dependent upon clarity of description. Three levels of increasingly detailed account were provided by ETICA concerning the findings regarding technology identification. First, schema in the forms of a simple list of the technologies and a diagrammatic representation of the technology descriptions (the *analytic grid*). Second, the technology descriptions themselves, typically around 2,000 words, provide more detailed information across a range of dimensions. Finally the many procedural and methodological accounts offer a deeper understanding of the technology descriptions themselves and provide first steps towards critical assessment. Accordingly, the technology findings are clearly stated and layered for a variety of audiences, enhancing the receptive potential of these findings for the project's main aim of policy development, but also facilitating the potential of the findings as a platform for further research. This is discussed in more detail below (*see How valuable is the research? p. 68*).

11. Is there adequate discussion of the study limitations?

As indicated above, ETICA documents discuss the limitations of the data sources analysed in terms of language and geography. However, there is less concern with the more procedural limitations we have identified in this assessment, but which are not identified as limitations within the ETICA documentation. In particular, no discussion occurs regarding the role of non-institutional actors in ICT innovation, the limited investigation of social factors at the level of the individual technology, or the social aspects of socio-technical innovation.

12. How valuable is the research?

This question is more relevant to assessment of an entire research project, rather than merely the first of three phases. However, we consider the development of the ICT taxonomy to be of benefit for foresight research in general, not just as a step within the ETICA project. It is not our position that the taxonomy is sufficiently robust to warrant use as a set of accepted premises upon which other research could be based. In our view additional work is required to enhance the social aspects of the technologies. Additional bibliometric analysis of the types indicated above (such as semantics) could add further depth to the understanding of the relationships between the technologies and issues, and could also be used in developing more detailed social analysis. Examining wider ranges of source material, including those from the social imagination, could lead to some restructuring of the taxonomy. In terms of using the taxonomy as a basis for foresight projections, we believe it would need to be allied with a more detailed analysis of innovation pathways which take greater account of disruptive innovations and non-institutional actors and of the interplay between innovation, adoption and power structures. However, these opportunities also indicate the value of the work already accomplished in that it provides a solid basis for more refined work, serving as a significant, if not complete, platform.

Even if the results themselves are not accepted, this phase of the ETICA project produced some valuable methodological innovations which can be used by others. The methodological procedures are documented to a sufficient degree that they can be used without inordinate adaptation to individual research goals. Here the extensive documentation of conceptual and procedural problems and their resolution processes is especially valuable. Such documentation aids the understanding of ETICA's methodologies, while also enabling one to adapt the methodologies if their underlying logic does not suit.

3.7. Conclusions

It is clear the ETICA project was a large and complex research project. The technology identification phase alone constituted a considerable project in its own right and made a valuable contribution to foresight research generally, both in terms of findings and development of foresight studies methodologies. The taxonomy of technologies and the analytic grid constitute significant conceptual developments of value to future researchers. Even if one discounts the valuable data with which ETICA populated these schemas, they remain important systematising tools with which to organise data. The use of meta-vignettes is an open-ended development which allows others to add further meta-vignettes to the existing set. This may allow enhanced and more rounded descriptions of the technologies, but it also risks the danger of discourse falling to the level of simply creating competing vignettes in order to make one's argument. There is also the danger that subsequent vignettes may lose their "meta" status and degrade to artefact descriptions. This suggests an need for further development of the meta-level analysis of the technologies themselves as a means of constraining vignette development.

As indicated at various points in the preceding discussion, much of what has been accomplished offers profitable opportunities for further research building upon the progress ETICA achieved. This does not require an uncritical acceptance of ETICA's work. Indeed, the most profitable avenues for further research stem from a critical analysis of ETICA's methodologies. The so-called "gaps" in the project represent research possibilities whose scopes and methodologies are delineated by the context of what *is* present, both in methodology and findings.

The key weaknesses of the technology identification phase of the ETICA project are likely to have only limited impact on the overall technology taxonomy, but have an undetermined effect on specific technology descriptions. The data structures did not give equal representation to the social elements of which technology was held to be composed, lacking elements for recording social data. The second potential weakness concerns possible researcher bias. We have seen that at least one corrective mechanism, the bibliometric data, does not appear to have functioned as designed. Furthermore, an apparent acceptance of normative framing by an indeterminate proportion of the researchers offers a vector for the introduction of bias and thus increases the possibility of its occurrence.

The key strengths stem from the scope of what was attempted. In attempting to cover the complete range of emerging ICT's and in drawing on such a large volume of sample data, the project was forced to develop strong research processes. Particularly noteworthy is the consistent attention to data validation at each step of the data processing through the cross-referencing against different sample formats via alternative research methodologies. The scale, range and depth of published findings provides valuable material for a wide audience.

In overall terms the technology identification phase can regarded as very successful; it produced findings in accord with the research goals, developed a strong framework for combining multiple foresight studies methodologies into a coherent research framework and followed those methodologies with only minor deviations. The resulting technology descriptions and taxonomy offer both conceptual insights and reusable frameworks. The issues we have identified as weaknesses do not significantly degrade the value of the findings at this stage of the ETICA process. However, they left our research in need of a description of the social aspects of technology as a socio-technical system.

3.8. Next Steps – Adding the social to ETICA's technology findings and answering a secondary research question.

The two papers published in 2015 in *Computers & Society*[10] answer the secondary research question:

"By what mechanism can an aspect of an emerging ICT have a restrictive effect on human autonomy?"

These papers presented the results of our efforts to provide a social account of ICT's and so fill the gap we identified in ETICA's findings. They also provided a generic characterisation of what it is to have human autonomy limited by ICT's which was drawn from contemporary analysis of current practice.

The purpose in producing such an analysis was to provide an account which could be used when analysing the future state of affairs which is the subject of the main research question. During this phase of the research we sought to analyse current circumstances from both a marxist and a capitalist perspective. The aim in so doing was to ensure the overall research project was not confined to either perspective, but offered findings which could be accepted under both. Additionally, our research indicated each had something of value to offer, and that the two complemented each other by focusing on different aspects of the current situation. Furthermore, both arrive at the same concerns, though they identify different causes and differ as to why these issues are worrisome. That both forms of analysis identified the same features as being problematic strengthens the acceptability of our findings.

Digital Alienation as the Foundation of Online Privacy Concerns (Dainow 2015a) developed a description of our current digital environment in the terms of existing marxist scholarship, mainly from within Europe. It extended the marxist concept of alienation into a uniquely "digital" form by revising Marx's analysis of the causes of alienation by updating it for the internet. By accounting for the mechanisms by which alienation occurs online today we developed a formal description of what it is to live in an autonomy-limiting digital environment and the mechanisms thereof. In doing so, this paper also outlines the relation between affordances, symbolic capital and other social dynamics so as to answer the secondary research question regarding how technical

[10] "Key Dialectics in Cloud Services" and "Digital Alienation as the Foundation of Online Privacy Concerns." *Computers & Society ETHICOMP Special Issue* (2015): 52-59 and 109–17

characteristics can cause psychological effects as a result of design and operational characteristics.

Key Dialectics in Cloud Computing (Dainow 2015b) provided an analysis of the current circumstances from a capitalist perspective, mainly from US scholarship. It focused on important debates within ICT ethics concerning current circumstances. This analysis provided an understanding of the mechanisms whereby the current capitalist model of digital services can cause ethically negative outcomes by focusing on specific issues. It also provided a model of some of the more specific and contingent social aspects of technology than possible in the more over-arching marxist analysis. In particular, we addressed issues of privacy, ethical responsibility, system architecture and business model, all of which have been shown to be influential in determining the impact the social dimensions of an ICT have on human autonomy. *Key Dialectics* therefore provided an answer to the secondary research question with regard to how elements such as business model and regulatory environment can restrict autonomy.

When considering the future in subsequent stages of the research, we projected the features described in the two papers onto the future state of affairs as demonstrably extant mechanisms with the potential to restrict autonomy. There may be other mechanisms and new ones may appear in the future. However, those we identified are sufficient in themselves to support restrictions of autonomy. This is demonstrated by the manner in which they generate ethical concerns today, as discussed in the papers. *Key Dialectics* also addressed the issue of the technical feasibility of doing things differently by detailing formal approaches and already-deployed systems which offer viable alternatives to those ways of providing digital services which we considered problematic.

The analyses in these two papers were used from this point forward in the research. The features identified in the two papers were subsequently imposed upon ETICA's technology groups as the model for how each technology may be deployed within the society in such a manner as to reduce autonomy. This is not to say they are the only features which could so restrict autonomy, but they are worthy of attention because we had demonstrated they are present today. We assume they will continue to be present into the future unless factors arise to address them, such as regulations. It is worth noting here that new developments such as GDPR do address some of the issues we identified.

4. The Current State of Affairs – Poles of the Debate

4.1. Introduction to the paper

The publication presented in this chapter seeks to identify key elements of the debate regarding ethical dimensions of ICT's relevant to autonomy. It forms part of the attempt to address the secondary question of "by what mechanisms can emerging ICT's threaten human autonomy?" We sought to answer this secondary question from both a marxist and a capitalist perspective. This paper uses the capitalist perspective, while the following chapter provides the marxist approach. The accounts in these two chapters form the foundation of the analysis addressing the primary research question.

The paper herein identifies three central dialectics within digital service delivery. These constitute defining positions regarding data privacy, ethical responsibility, technical architecture and economics. These constitute the main frameworks within which ethical discussions of digital services occur. While these are framed in terms of cloud computing, they are applicable to all systems which deliver personalised services to humans. At present, such systems are almost exclusively based on a client-server architecture, in which the server occupies a virtualised cloud layer (Moglen 2010). Such an architecture is our primary ethical concern regarding the socio-economic factors relevant within ICT's because it places the data and algorithms outside the knowledge and control of the individual. While it may be possible to restrict autonomy with purely client-based systems, it is significantly harder to do so. Client-only systems grant control over data access and use to the individual concerned (Ng and Wakenshaw 2015). Since the economic success of many online services has been dependent upon a client-server architecture, we anticipate considerable effort to maintain a client-server architecture on the part of major players in the digital industries, such as Facebook and Google.

The rest of this chapter contains the paper as published. Minor modifications have been made to ensure the writing style matches that of this thesis (such as changing "I conclude" to "we conclude"). The citation style has been changed to maintain consistency with the rest of the thesis. In keeping with the need to present the published material in the format of a coherent sequence of chapters, journal meta-information, such as keywords and categories, have not been reproduced here. The original publication did not number headings. Numbering has been added to maintain consistency with the rest of the thesis.

4.2. Key Dialectics in Cloud Services

Computers & Society ETHICOMP Special Issue (2015): 52–59. https://doi.org/10.1145/2874239.2874247.

4.2.1. Introduction

This paper explores dialectics within debates over key ethical issues pertaining to cloud services. These issues concern privacy, responsibility for the actions of systems, and the development of monopoly service providers. Between them these concerns largely dictate the shape and capabilities of current and future cloud-based services. We shall show how the current state of affairs is dominated by a sense of lack of agency in terms of doing things differently from the current reflexive practice, an assumption that no alternatives to current practice are possible. This paper will attempt to organise the key concerns with cloud services by organising them into three dialectical axes:

- The nature of the relationship between personal privacy and service provision.
- The degree to which people who build or operate cloud-based services are ethically responsible for the actions or effects of those services.
- The nature of the marketplace for those services.

Since this chapter considers cloud services in the broadest sense, it is appropriate to commence with the definition of cloud computing used here. The *US National Institute of Standards and Technology Special Publication 800-145* defines the essential characteristics of cloud computing as being:

- The ability to provide services whenever desired without human intervention.
- Being available to a wide range of client devices via networking technology.
- The "virtualisation" of computing resources, such that digital operations are not linked to specific servers or locations.
- Scalability – the capability of the systems to scale up or down in response to changes in demand (a necessary corollary of virtualisation).
- Often, but not necessarily, Software as a Service.

(Mell and Grance 2011)

Clearly this definition applies to many, if not most, internet systems and digital services, not merely to the virtualisation of server functions previously found in the traditional client-server network. Under this view, Facebook and Google's search engine are both cloud services. This is both valid and important - confining discussion of cloud computing to data processing or file storage functions limits discussion to a few contingent uses of a wider system and obscures the essential factors we need to consider.

4.2.2. Privacy versus Security

Our first axis is the necessity versus the contingency of reductions to privacy under new digital services. That is to say, there is one body of opinion which holds that the erosion of personal privacy is a necessary and unavoidable consequence of, or precondition for, the delivery of digital services. These positions tend to be a reflexive response within the development community, rarely stated formally, and is a minority view in the literature. As a result, detailed arguments as to why privacy *must* be reduced to enable cloud services are scarce. However, Professor Bergkamp's paper, *The Privacy Fallacy* (Bergkamp 2002) marshals all the arguments in this camp.

Professor Bergkamp argues there should be no privacy protection of any form because preservation of personal privacy is harmful to society in many ways. He provides five main arguments; there is no need for data privacy, data protection reduces individual freedom, personal privacy is contrary to economic growth, EU data legislation is unenforceable and the EU's data protection regimes put it out of step with the rest of the planet. We will now explore each of these in more depth.

Bergkamp argues there is no need for data protection or digital privacy because no one wants it and it serves no purpose. He states there is no evidence anyone has ever been harmed by privacy violations or personalization of services based on personal data. He does not provide any evidence for this and it is contradictory to the reported activity of many data protection authorities. For example, in 2014 the Data Commissioner of Ireland received 2,264 data breach notifications, investigated 960 complaints and launched 162 prosecutions. Half (53%) of complaints involved disclosing personal data inappropriately, such as disclosure of personal financial data to relatives or the listing of email addresses and passwords on public websites (Office of the Data Protection Commissioner 2015). The *Verizon 2014 Data Breach Investigations Report* (Verizon 2014) covers 63,000 data violations across 93 countries in 2014. It highlights financial theft and the cost of dealing with a breach, such as cancelling credit cards, as the main harms to the individual. Other research exists to show harm from less obvious privacy violations. *RT@Iwantprivacy: Widespread violation of privacy settings in the Twitter social network* (Meeder et al. 2010) details harm from privacy violations in Twitter when people reuse private tweets in public. *Privacy Violations Using Microtargeted Ads* (Korolova 2011) examines harm from privacy violations in Facebook. Privacy violations has also been shown to harm the companies themselves. *How Privacy Flaws Affect Consumer Perception* (Afroz et al. 2013) shows how privacy breaches reduce the chance people will buy from a company, while *Is There a Cost to Privacy Breaches?* (Acquisti, Friedman, and Telang 2006) shows how privacy violations reduce a company's share price. Studies also exist to show harm from personalization of advertising and news. The research findings of Sweeney's *Discrimination in Online Ad Delivery* (Sweeney 2013) show how racial stereotyping in ad personalization harms Afro-Americans in many ways, including job prospects and access to financial services. *Bursting Your Filter Bubble* (Resnick et al. 2013) shows harm from news personalization, while the famous Facebook news manipulation study, *Experimental Evidence of Massive-scale Emotional Contagion through Social Networks* (Kramer, Guillory, and Hancock 2014) shows how personalizing news feeds to contain more negative contents can depress people. Bergkamp's proposition that there has never been any harm from privacy violations or personalization appears to be contradicted by such evidence.

Bergkamp also argues there is no need for data protection because no one wants it. He argues that people don't realise that data protection prevents personalization, but that when they do, they always prefer personalization over data protection. He does not cite any evidence for this. By contrast, Culnan's 1993 study of personalization in shopping, *How Did They Get My Name?* (Culnan 1993) shows that when offered the choice, the people he surveyed preferred privacy over personalisation. More recently, the *2013 Comres Big Brother Watch Survey* (Comres 2013) polled 10,000 people in nine EU countries to find 75% were concerned about privacy and wanted data protection regulations, while 45% believed they were being harmed by corporate data practices.

Bergkamp also argues there is no need to regulate sale of personal data because companies never sell it. However, there is, in fact, a huge industry in the sale and aggregation of personal data, as the 2014 Federal Trade Commission's investigation into data brokers found (Brooks 2005; Federal Trade Commission 2014).

Bergkamp argues that personalization results in cheaper prices. However, he does not cite any empirical evidence for this or reasons why it should be so. He cites as evidence a statement made by Fred Cate, Professor of Law at Indiana University, that personalization results in cheaper prices, but this was a statement made to a Congressional committee, not a research finding. Prof. Cate's own list of publications does not include any research into personalization; his speciality is data protection law. Later in the paper Bergkamp states that data protection costs money and that it is so burdensome and expensive that businesses can only survive by ignoring their legal obligations. One may imagine he believes this is the cause of higher prices to consumers, though he does not say so. However, research like Sweeney's *Discrimination in Online Ad Delivery* (Sweeney 2013) shows how personalization actually increases costs to Afro-American consumers in the USA, while Turow's *The Daily You: How the New Advertising Industry is Defining Your Identity and Your World* (Turow 2011) shows how personalization can reduce or increase prices, depending on whether you are the consumer companies want or not.

Bergkamp also argues that privacy protection increases identity theft because data protection makes it harder to tell if someone really is who they claim to be. He does not cite any evidence for this and it seems counter-intuitive. Given that privacy protection reduces access to the personal data necessary for identity theft, such protection could be presumed to make it harder to commit, so one could argue the exact opposite of Bergkamp in the absence of any research. Bergkamp's position here allies with his arguments elsewhere in his paper that we all need to know as much as possible about each other in order to protect ourselves from one another and that privacy directly prevents this. He states that one problem with privacy protection is that it allows an individual to control what they disclose to the world. He does not explicitly say this is a bad thing, but it is clearly implied from his usage. Here it is worth noting research showing the reverse, that lack of privacy restricts human freedom. For example, knowledge one is being watched on the internet has been shown to have a chilling effect on what people say (Bass 2013) and what they search for (Marthews and Tucker 2014), even when engaging in legal and socially acceptable activity.

Bergkamp claims there is a vast amount of money to be made acquiring and selling personal data, despite his earlier claim that there are no businesses selling it. He provides no evidence for this economic activity, but the claim is supported elsewhere. For example, in 2013 the OECD estimated the personal data of each Facebook user to range from $US40/year to $US400 (OECD 2013b). Bergkamp claims that this data market alone is sufficient reason to remove privacy protections. However, the mere presence of economic activity does not, in and of itself, mean we should encourage it. There is a vast amount of money to be made in drug smuggling, but no one uses that as an argument for encouraging it.

Bergkamp also states that the EU's data legislation is unenforceable. He says the very concept of personal privacy is too vague to support regulation and that the regulations cannot properly specify what constitutes personal data. Furthermore, he says, each privacy incident must be judged on its own merits. He does not explain how judging a case on its own merits is a problem. Each and every infraction of the law is judged individually, so arguing that this is also the case for privacy issues does not, in and of itself, constitute a sign of poor legislation. Furthermore, it is difficult to imagine what the alternative would be if a regulator or judge was not allowed to consider the specific details of each case they were trying to adjudicate.

As stated earlier, Bergkamp believes that data protection is so onerous that no business can do it properly and survive financially. He claims the only outcome is that data regulations are never enforced. Clearly the many cases of prosecution for privacy violations are not accounted for in this argument. Bergkamp also states that EU data protection legislation is founded on a misunderstanding of how business works, but does not provide any further details regarding the nature of the misunderstanding or what the reality truly is, so this statement is impossible to assess.

Bergkamp's paper also states privacy protection damages society because it involves the government paternalistically interfering in people's relations with each other in a misguided attempt to stop people hurting each other. Such an argument can also be said of laws against violence and theft, so the logical consequence of such a position is that we should move to a state of complete anarchy. However, Bergkamp does not address this implication. Instead he goes on to state that government's should never restrict any information under any circumstances. Again no reasons are provided to justify this proposition. Such a broad statement can also be used as an argument in favour of making child pornography freely available, so some additional clarification would seem appropriate.

Finally, Bergkamp claims that EU data protection legislation is out of step with the rest of the world. He does not provide any evidence to support this, but he clearly thinks this is a bad thing and grounds for abandoning data protection. This is a questionable claim. The EU's data protection regime was intentionally built to accord with pre-existing OECD guidelines, which were first developed in 1980 (OECD 2013a).

Bergkamp never states it is technically impossible to maintain privacy while extending cloud services. His arguments are merely that we should not. Our position is that there is no necessary and unavoidable relationship between privacy consequences and functionality. One does not *have* to reduce privacy in order to extend services. Rather, it is always a question of choice, either in how the system is constructed or in the type of business model under which it operates, and there are *always* alternatives. It may be that some of those alternatives are more expensive than the privacy-reducing models, or that alternatives are more technically challenging. However, that, in and of itself, is not an argument for the necessity of privacy-reducing models, but rather an argument underpinning a particular business model or software approach.

Currently those who are building cloud-based services most commonly work on the basis that privacy is exchanged for digital services. However, there is also a growing body of those seeking to develop alternatives, in terms of governance or business model or in terms of code. The most notable is the Privacy by Design movement. However, most of the Privacy by Design material is so vague as to be little more than statements of intent. For example, IBM claim to have moved to Privacy by Design by doing nothing more than implementing awareness training and building an internal system for reporting data breaches (Office of the Information & Privacy Commissioner of Ontario and IBM 2011). Here Langheinrich's paper, *Principles of Privacy-Aware Ubiquitous Systems* (Langheinrich 2001) stands out as the exception, being a concrete statement of specific technical design principles which genuinely do embed privacy considerations into the technical architecture. Langheinrich's paper shows it is possible to build robust systems which have privacy protection embedded within the design and operation of the system.

It is notable that Langheinrich has practical experience in the design of privacy systems, being one of the authors of the W3C's technical standard, *Platform for Privacy Preferences*, or PPP (Cranor et al. 2002). The PPP standard enables browsers to hold the user's preferences for what data they will allow a website to gather. The server component of PPP allows the web server to list its own data-gathering practices. PPP then enables the browser to compare the web server's practices with the user's preferences. The system provides for warnings to the user and for compact and rapid communication between client and server of data practices. The system was supported in Microsoft Internet Explorer 6 when it first emerged, but lack of support by website owners means PPP is largely unused today.

The first of Langheinrich's principles is the Principle of Openness, or "Notice." This simply states that no device or service should gather data about someone without telling them. Here he makes reference to PPP as providing a digital vocabulary which could be used to programmatically describe what data is being gathered, for what purpose and by whom. This is paired with the second principle, the Principle of Consent, which encodes the legal necessity for informed consent. A system must allow for someone to opt out of being tracked or recorded, and do so without denying service on a "take it or leave it" basis. Thus, for example, buildings would need to disable tracking for some people and not simply refuse them entry.

The third principle is termed "Anonymity and Pseudonymity." This states that people must have the option to remain anonymous. The issue here is that some services are only possible if they know a user's identity and history. Here Langheinrich introduces pseudonymity. Under this system a person may have a unique identifier of some form, such as a cookie or RFID chip, which anchors the data systems and forms the index key to their personal data history. However, this identifier contains no personally identifiable information and is discardable at any time. Furthermore, such a system permits people to have multiple pseudonymous ID's and so prevent aggregation of disparate activities by data brokers. It is noteworthy that EU data regulations have recently been updated to add the category of pseudonymous identity between personal and anonymous data (Ustaran and Macmillan 2015).

Langheinrich's fourth principle of "Proximity and Locality" limits the scope of data collection. Looking to a future in which people have many devices capable of recording their surroundings, the principle of proximity states that these devices can only operate in the proximity of their owner. This prevents people leaving devices to record data unseen then returning for them later. Of wider application is the principle of locality; devices should not transmit data any further than absolutely necessary to fulfil their functions. For example, Samsung's voice activated TV's transmit all conversations they hear to Samsung's central servers. Voice commands are interpreted there and the appropriate command then sent back to the TV. All conversation recordings are stored permanently for later analysis ('Privacy | Samsung UK' n.d.). Under Langheinrich's principles the TV would have been designed so that it did not need to involve cloud services. Voice recognition chips have been around for 20 years and could easily have be used instead.

The fifth principle is the "Need for Security," in which Langheinrich advocates various levels of security depending on the nature of the data. More importantly, he illustrates how the previous principles themselves enhance security. If data is not being transmitted many security problems simply vanish. Similarly, if data is not linked to an identifiable individual, but only to a pseudonymous ID, unauthorised access has less potential for harm.

Langheinrich's final principles are the principles of "Collection and Use Limitation." These state that data collectors should only collect data for a specific purpose and not store it, as Samsung TV does, in case they want to use it in the future. Secondly, they should only collect the data they need in order to fulfil their task and nothing more. Finally, they should only keep data as long as it is necessary for the purpose. While these appear primarily legislative principles, they can be embodied in technical design through the use of the earlier principles. For example, if data is housed in the user's devices in accordance with the principle of locality, then the user can impose usage and storage limitations themselves.

Langheinrich's principles, if implemented, would solve many privacy concerns, enhance security and actually make many applications of ubiquitous and cloud services easier to construct. What they show is that it is perfectly possible to design cloud services in a manner which enhances both security and privacy at the same time, while permitting all the personalization necessary. They place control of personal data firmly in the hands of the user without compromising technical operations in any way. In fact, their reduced dependence on permanent access to centralised services makes them more robust and reduces the burden of traffic on the internet. These principles are easy to understand and yet produce powerful architectures. They offer a practical and detailed response to the reflexive position that personalized cloud services must reduce privacy. In doing so they provide concrete evidence that it would be possible to move cloud service evolution into a path which fulfils all its potential, yet enhances privacy and security at the same time. Langheinrich's design principles demonstrate that the reduction of privacy in cloud services is a choice, not a necessity.

4.2.3. Ethical Responsibility

Our second axis is concerned with the degree to which people who design, build or operate cloud services are ethically responsible for the consequences of the actions of those systems. This question does not arise with regard to all cloud services, but only with the rising generation of autonomous services which process personal data in order to deliver personalised services, such as personalised search results, product recommendations and news feeds. In the near future we will see the rise of more intelligent and more life-critical personalised services, most notably with bio-implantation and other medical services (Ikonen et al. 2010). The question is primarily one of who is responsible when such autonomous services make decisions which result in harm, but where these decisions are not the result of faulty design or incorrect data.

The competing positions are that, on the one hand, programmers and operators are not ethically responsible for the actions of autonomous systems, versus a view that they are. It is difficult to argue that the person holding a hammer is not ethically responsible for the consequences of whatever happens when the hammer hits something because the hammer is totally under the control of the user. However, with large industrially-produced complex automated systems, especially those that include some form of AI functionality, arguments emerge in favour of the position that those who build the systems are not ethically responsible for the decisions those systems make. This argument will no doubt be exacerbated the more powerful and the more intelligent and autonomous these systems become. This issue is discussed most frequently with regard to autonomous military systems, whose lethality makes the question of ethical responsibility both stark and urgent. However, the question is just as pertinent for any form of autonomous system, including those cloud-based personalization systems already in operation.

Andreas Matthias' paper, *The Responsibility Gap* (Matthias 2004) offers a fairly straightforward account of the philosophical logic behind the position that programmers are *not* responsible for the actions of their autonomous systems, while Robert Sparrow's *Killer Robots* (Sparrow 2007) presents the same view via an examination of the practicalities of creating and deploying autonomous systems. Both take the position that no one at all is ethically responsible for the actions of autonomous agents.

Sparrow's argument is based on his particular understandings of the terms 'autonomy' and 'responsibility.' Our view is that he defines these terms in such a way as to make any contrary conclusion impossible. Early in the paper, he defines autonomy as being free from external causation:

> "Where an agent acts autonomously, then, it is not possible to hold anyone else responsible for its actions. In so far as the agent's actions were its own and stemmed from its own ends, others cannot be held responsible for them. Conversely, if we hold anyone else responsible for the actions of an agent, we must hold that, in relation to those acts at least, they were not autonomous." (Sparrow 2007, 65–66)

Sparrow does not defend this definition of autonomy. However, once it has been defined this way, it becomes a matter of logical necessity that there is no ethical responsibility by the programmers or controllers. It is also worth noting that Sparrow uses autonomy in an absolute sense, as if the agent were free from all influence except their prior experience. In particular, he does not recognise the environment, the capabilities of the device or its internal structures as having any impact on decision-making. He argues the programmer cannot be responsible because the essence of an autonomous system is that it will make unpredictable decisions. He argues the controller of the system is not responsible because they could not anticipate what it would do any better than the programmer. In both cases, he ignores the fact the system is designed to perform a particular role in a particular environment. A software agent is not free to do just anything, it can only recognise inputs of a type it has been designed for, and has a relatively limited range of actions it can take, and can only operate in a specific type of environment. A share-dealing system cannot walk the dog or assess your exercise regime. The type of decisions an autonomous system may make, and the range of options available to it, are not only predictable, they are the basis upon which it was designed and built - they define it. An autonomous system may make its own decisions, even alter its own programming, but its range of actions and the forms of harm it may commit are knowable in advance in virtue of the type of system it is.

Sparrow does not mention Strawson in his paper, but his conception of moral responsibility has close parallels to Strawson's influential work. Strawson's position is that no one is morally responsible for anything because no one is free from external influence (Strawson 1994), though the details of why are beyond the scope of this research project. Though Sparrow does not say so explicitly, his use of responsibility is clearly that one can only be responsible for specific actions. In Sparrow's view, the design of the system and the decision to use it do not carry any ethical responsibility because neither gives one the ability to predict the specifics of an individual act the system may take.

Sparrow also argues it is not possible to hold the system itself responsible because responsibility necessarily requires punishability which requires suffering. Under his definitions, something can only be morally responsible if it can be punished and something can only be punished if it can suffer. Since software systems cannot be made to suffer, they cannot be punished and so cannot be held responsible for their actions. Note here that we have switched from talk of "being responsible" to talk of "being held responsible." This demonstrates that Sparrow has conflated the moral state of being responsible with the social status of being eligible for punishment.

Matthias's *The Responsibility Gap* (Matthias 2004) also argues that no one is responsible for the decisions of autonomous software systems. His position also links responsibility to individual acts, holding that one can only be responsible if one can know the internal state of the system *and* has control of each act it takes, at least to the degree where one could prevent it. Under this analysis a programmer has no responsibility for the actions of a system once the owner takes control. The owner is not responsible because they cannot know the internal state of the system. Matthias spends some time examining different types of AI learning, showing how each makes their internal state unknowable in different ways, but the differences do not affect his final conclusion.

The narrow understanding of responsibility seen in Matthias and Sparrow is the foundation on which their arguments rest. In contrast, Miller's *Collective Responsibility and Information and Communication Technology* (S. Miller 2008) confronts this issue by arguing there are different types of responsibility. In addition to the responsibility for individual acts which Matthias and Sparrow focus on, Miller points out we also recognise one can have "structural" responsibility by creating the conditions which made the act possible or by ordering others to take actions which eventually led to the act. Under Miller's analysis both programmers and controllers of autonomous systems take structural responsibility for every act taken by these systems. Miller then goes deeper, investigating the concept of collective responsibility. He argues that to the degree that individuals contribute something to the shape and operation of an autonomous system, so they share in responsibility for its actions. Here he acknowledges the existence of corporate responsibility, but argues that it does not provide a moral shield for the individual workers, whose individual contributions to a system's operation convey a share in collective responsibility for its actions.

Miller's approach is a step towards recognition that software agents exist within the wider context of human activity. This broader perspective is fully achieved in *Software Agents, Anticipatory Ethics, and Accountability* (D. G. Johnson 2011). Reiterating the perspective that technology is socially situated, this paper argues that the concept of any digital service as an autonomous agent is merely metaphorical; that no such system can be autonomous in the sense we apply autonomy to humans in moral debates. As such, the use of the metaphor is justifiable only by its utility. Johnson criticises the concept of any software as an autonomous agent on the grounds it generates just these ethical problems. Instead, Johnson argues we should recognise autonomous systems as elements within a larger socio-technical system, made by people and used by people for human purposes. Under this view autonomous systems are not independent entities hermetically sealed from their environments, but systems which can only be understood by reference to the context of their use. Johnson makes implied use of the different forms of responsibility seen in Miller, but does not elucidate the differences. Instead she focuses on the arbitrariness of delimiting technical artefacts. She denies that autonomous software agents are different in kind from any other form of automated or semi-automated device, being merely more complicated. Autonomous systems are thus merely, like hammers, extensions of human will and intent. Under this arrangement, ethical responsibility for their actions is not in any way changed by the mere fact of their complexity.

4.2.4. Open versus Closed

Our final dialectic concerns the form and marketplace of cloud services. Here the dominating dialectic is that of open versus closed systems and open versus closed organisational contexts for such systems.

The scene for this debate is best set by Eben Moglen in his presentation, *Freedom in the Cloud*, delivered to the Internet Society in 2010 (Moglen 2010). Moglen argues that the internet was originally designed as a non-hierarchical peer-to-peer network. However, under the influence of the architectural model of client-server networking, the services which evolved used a smart-server-dumb-client model, in which both algorithms and data were centralised. Moglen maintains that cloud architecture works on this thin client - fat server model and does not represent a new computing architecture, merely the virtualisation of some server operations within this traditional model. These servers maintain activity logs. These logs can be mined for behavioural data. Marketing companies learned they could mine these logs to understand, predict and influence user behaviour in order to sell advertising. Moglen contends that as the perceived value of this information grew, it spurred the development of a secondary internet infrastructure of tracking services designed to add to the growing database of what we now call *user profiles*.

Thus, Moglen describes how an architecture which concentrates processing power and data at centralised locations promotes a concentration of both technical proficiency and economic power, while also promoting a top-down hierarchical organisational model and, in a global internet, the development of a limited number of very large monopoly service providers. This has, he argues, produced an extreme power dichotomy between those who own the services and those who use them. The business model which has come to dominate the internet is that of delivering services in exchange for spying on the users all the time. Moglen describes this state of affairs as undesirable for two reasons. Firstly, the price is too high and the services are not worth the loss of privacy. Secondly, the lack of alternative models for access to the same services makes this unfair arrangement unavoidable. He argues that we need an alternative architecture in which the data about us stored on centralised servers is instead housed in devices we own and carry with us. We can then control who accesses this data and how. He argues that this is possible with current technology.

There is an additional element of concern within Moglen's model which he hints at but does not explore. The combination of architecture and business model he describes has produced *walled gardens*. These are silos of private technology and proprietary data formats which are not compatible with, or accessible by, other systems or organisations. The patent system combines with a capitalist marketplace to financially reward such behaviour. If I am the sole owner of a system everyone wants to use, I can make money. If I create a system which I give away, I do not benefit. What I therefore need to do is lock everyone into my technology, and then I will "lock in" the market (Crampton and Boudreaux 2003).

The effect of this is to lock data and services into a single monopoly provider. The provider becomes the gatekeeper over the knowledge of what they do and how they do it. Users cannot migrate to a competitor without significant effort and loss. For example, if you close your account with Amazon, they will remove all the books from your Kindle (Amazon 2017). You cannot therefore switch to an alternative, such as Adobe Digital Editions, without re-purchasing your entire digital library. Different legal regimes permit different levels of access inside these walled gardens, but in no case does a society have full knowledge or any substantive control. Such a system has no interest in open standards, interoperability, or a free flow of information. This lack of interoperability and open standards was why the internet and HTML were not developed by commercial enterprises. Early pre-cursors of the web tried the same walled garden approach, including America Online, CompuServe and Lotus Notes. It was only when Tim Berners-Lee gave HTML away that we broke free of this limiting system and gained the web. Berners-Lee gave it away because he saw things in exactly this way and believed that if he patented or sold HTML, it would become just another walled garden (Berners-Lee 1999).

However, as companies have developed services which sit atop these communally-owned standards, so they have developed further proprietary systems. The final result is that companies have built a new layer of walled gardens and data silos on top of the open platform which is the internet (Dyhouse 2010). The scale of the internet user base combines with a shared service delivery infrastructure to enable the rise of extremely large global monopolies, such Google, Amazon and Facebook. The result is that cloud services are portioned out amongst a limited number of very large hierarchical organisations, each of which hides its use of data from public scrutiny and uses its monopoly position and ownership of personal data as a competitive advantage (McChesney 2014; Curran, Fenton, and Freedman 2012; Brown and Marsden 2013a). The net effect is that people are locked to service providers like serfs to their lord. However, unlike in the Middle Ages, there is no competing lord to flee to if you are unhappy with your lot. This power is a concern to many. Some argue, for example, that Amazon's potential to control what books are available makes it a political institution as well as an economic one (Brown and Marsden 2013c), while Google is has consistently been one of the biggest spenders on lobbying in Washington since 2012 (Hamburger and Gold 2014).

Opposing this state of affairs are a disparate range of alternatives, such as Moglen and his concept of a personal server. Each alternative tends to focus on one aspect of this system, such as technical architecture or business model. Technically, the existence of the internet is based on open standards, such as TCP and IP (Stevens and Wright 1994), so alternatives have always been available on a technical level. Here we have the open source activists, such as the Free Software Foundation and the IETF. In addition, we have less obvious alternative architectures based on peer-to-peer (as opposed to client-server) models, such as the BOINC platform for community computing (D. P. Anderson and Fedak 2006) and the BitTorrent protocol (Pouwelse et al. 2005). Standards like XML (XML Working Group 2015) and RDF (RDF Working Group 2015) provide a means of breaking open walled gardens through data exchange, while people such as Chris Marsden in the UK or Robert McChesney in the USA have developed the rationale for breaking down these proprietary data silos.

McChesney argues that the development of monopolies and cartels has so dominated the internet that there has been little economic benefit for the rest of society. He argues that there is so little competition at the point of delivery that service providers constitute a cartel which should be forced into competition with not-for-profit public alternatives. He calls for the monopolistic corporations dominating important services, like Facebook and Google, to be broken into smaller competing units and subject to much more stringent and detailed state control. McChesney's argument is that the size of these corporations is so great they pose a threat to democracy itself through their power to lobby politicians, dominate online debate and skew economic development (McChesney 2014).

Concern over monopoly domination is addressed in a different manner by Brown and Marsden in a number of publications. Instead of seeking a solution by changing the economic structure, they focus on the proprietary data structures which form the foundation of such domination. In addition to rights such as the right to have one's records deleted, they argue for the right to move such data to an alternative provider of the same service (Brown and Marsden 2013b). They cite similar historical examples in which Microsoft, IBM and Intel have been forced into making their systems interoperable with competitors, mainly through antitrust approaches in the USA and EU (Brown and Marsden 2013a). They argue that state intervention to break up these monopolies is not practical in a world dominated by competing national legislative regimes. Instead, they argue that merely providing users the ability to switch to alternatives would be sufficient. They believe that this would stimulate the development of service providers offering a range of alternative models (Brown and Marsden 2013c).

The approach of treating monopoly service providers as public utilities is gaining ground in government circles. Recently the UK House of Lords called for the internet to be treated like a public utility rather than a market place of optional luxuries (The Select Committee on Digital Skills 2015). International bodies, such as the EU and UNESCO, have started calling for wider civic involvement in determining how services are provided (UNESCO Secretariat 2014; Kroes and Buhr 2014) and for the development of alternative service provision models. For example, the outgoing EU Vice President, Neelie Kroes, stated in November 2014:

> "Why should we have to give up our privacy for a "free" service if we prefer to pay for that same service with cash and keep our privacy?" (Kroes and Buhr 2014, 1)

4.2.5. Conclusions - Agency

While these three dialectics focus on different issues, the poles of each axis rest on competing perspectives on the possibility of agency. Those who accept things as they are now do not see a possibility for agency, while their opponents do. On the first axis we have those who hold that preservation of privacy and delivery of service are necessarily in opposition. Here they are holding that there is no possibility of agency in the relationship between privacy and service design. To the contrary, we have seen how Langheinrich's design principles show multiple opportunities to intervene in the ways which deliver services while also maintaining privacy. In our second axis of ethical responsibility, the position of there being no ethical connection between the creator of an autonomous system and that system's effects is also a position of there being no agency. Here lack of agency pertains not to the nature of the system, but to the consequences of the system's actions. Under this view, once an autonomous system is activated, human agency ceases. However, as we have seen, preserving a lack of responsibility requires limiting the conception of where agency lies. Responsibility has to be defined in a very constricted manner which focuses on the making of each individual decision and denies the influence of any context. Instead, services are treated as independent of any human agency - in terms of their design, their environment, their purpose, how they are used and who benefits. By contrast, once autonomous services are contextualised within a field of human practice, human agency becomes apparent throughout the construction and operation of such systems and human ethical responsibility becomes self-evident. Finally, in our third axis of service architecture and business model, we see a historical lack of agency in the development of the broader internet culture. Here the client-server structure was accepted reflexively by developers and users, along with the most obvious reflections of this in organisational and economic models.

In all three debates we see one pole in each dialectic disempowering itself, primarily because it simply fails to recognise that there is a choice and that agency, the power to act differently, exists. The conclusion which emerges from this is that a key step to improving the current ethical status of cloud services is inculcating in programmers and leaders that they possess agency, bringing them to recognise there are alternatives and that they have the power to explore them.

- end of published paper -

4.3. Next Steps

Accepting the analysis of some of the dynamics and structures of the internet economy which are ethically problematic, the following chapter seeks to describe the overall impact it has on the individual. If these structures and dynamics continue as they are, they will provide the socio-economic structures by which emerging ICT's may restrict autonomy in the future. Having identified these structures and dynamics in *Key Dialectics*, the next task in the research process was to develop an account of how they can restrict human autonomy, to argue that they do so today, and to provide a framework for describing such a state. The following chapter addressed these issues. It was published as "Digital Alienation as the Foundation of Privacy Concerns" in *Computers & Society* ETHICOMP Special Issue (2015).

5. Digital Alienation - ICT's as Socio-Technical Systems

This chapter details our work to expand the social aspects of technologies as socio-technical systems and to account for mechanisms by which someone may be alienated from their digital environment. It forms the second part of the answer to the secondary research question of by what mechanisms can ICT's restrict human autonomy.

5.1. Background to the Paper

Our evaluation of ETICA's technology groups indicated the need for additional work to provide an account of the social dimensions of ICT's. In particular, there was a need to account for social aspects of ICT's which impacted human autonomy, especially those which were not dictated by the necessities of engineering. These are important because they indicate that it is technically feasible to achieve the same goal without the attendant ethical concerns. In order to have any practical application in computer science, a social account within computer science must take into consideration how the created output of the field, such as user interfaces or algorithms, can have the non-technical effect of restricting human autonomy, as was summarised above (see *2.5 Addressing the social aspects of technology, p.31*).

The following section contains the published paper addressing this issue, *Digital Alienation as the Foundation of Online Privacy Concerns* (Dainow 2015a). Accepting and building on the analysis in the preceding paper, it explored the difficulties of using a traditional marxist analysis to account for digital alienation. The problem for marxist analysis is that the activity people undertake online does not look coerced nor does it appear estranged from the creator's individuality, both of which are typically seen as necessary for the production of alienation. As a result of this apparent difficulty, much of the existing scholarship has focused on the relationship between alienation and labour on the internet. We overcame these difficulties by discarding the traditional approach to argue one can better understand alienation online by focusing on the relationship between user intent and technical infrastructure, rather than concerns with the nature of labour. We argued that under the existing economic model dominating the internet, free services are financed by recording user activity and then using the products of this commercial surveillance to sell information about people to others. The paper argued the real harm in current online business models is that commercial surveillance is being used to commodify private life. Seeking to define personal data in more precise terms, we introduced two new concepts necessary for a detailed discussion of any ethical issues regarding personal data - the *digital shadow* and the *digital persona*. The digital shadow is a relatively common concept taken from Alan Westin (Westin 1970). The digital persona is the term we invented to account for personal involvement in one's online communications. Here we defined a uniquely digital version of a concept known in Psychoanalysis as the 'persona' (Jung 1977) and in sociology as the 'mask' (Goffman 1990). We then show how affordances in current online systems are tuned to commodification of the user's personality. We then explore the nature of online surveillance and show how affordances combine with the surveillance economy to produce digital alienation.

The rest of this chapter contains the paper as published. Minor modifications have been made to ensure the writing style matches that of this thesis (such as changing "I conclude" to "we conclude"). The citation style has been changed to maintain consistency with the rest of the thesis. In keeping with the need to present the published material in the format of a coherent sequence of chapters, journal meta-information, such as keywords and categories, have not been reproduced here. The original publication did not number headings. Numbering has been added to maintain consistency with the rest of the thesis.

5.2. Digital Alienation as the Foundation of Online Privacy Concerns

Computers & Society ETHICOMP Special Issue (2015): 109–17. https://doi.org/10.1145/2874239.2874255.

Digital alienation is a privacy issue. Digital alienation occurs when one's digital lifeworld or the digital self is exploited. The process of exploitation extracts value from a person's digital activity through coercion and manipulation. We are coerced into submission to ubiquitous commercial surveillance[11] of our digital activity. Value is extracted from this surveillance process through the conversion of surveilled data into economic and political capital. The entire system represents a reification of one's digital lifeworld and commodification of the digital self. It also poses a number of problems for traditional understandings of concepts related to alienation within marxist theory, such as coercion, exploitation, and power dynamics. Indeed, much of the debate in this area over the last few years has been concerned with how to account for alienation within a digital context. We argue solutions to current problems can best be achieved by altering the analytic approach.

[11] Commercial surveillance involves the recording and analysis of online user behaviour with the aim of predicting and controlling their behaviour (Turow 2011).

5.2.1. Scope of Concern

Being connected to the ubiquitous computing environment which is coming to surround us is already necessary for full participation in modern Western societies. A review across the range of those emerging ICT's which will impact society over the next decade shows that being connected may become necessary for survival itself (Ikonen et al. 2010). When the internet first emerged, it was predicted that it would "flatten" the power structures of traditional society, even lead to the "fading away" of the nation state (Negroponte 1996). Such views were based on technological determinism; they envisioned the new distinguishing features of internet technology as passing unmodified into society and reshaping it to match the internet's technical architecture (Curran, Fenton, and Freedman 2012).

In reality, the development of the internet ecosystem has been filtered through the structures of pre-existing society and evolved in accordance with its imperatives. While it has been disruptive in terms of changing some of the dominant players in media markets, destroying some and creating others, it has not fundamentally changed the power structures in society. Authoritarian governments have learned to control and censor it, hegemonic corporate capitalism has come to dominate it, and people's digital activities have been cajoled into closed silos controlled by a very few exceptionally large corporations (Curran, Fenton, and Freedman 2012). Once seen as the antidote to structural inequality, the internet has actually become a profoundly powerful tool of domination based on exploitation and alienation.

5.2.2. Alienation in Critical IS Studies

The term 'digital alienation' is used in Critical IS research to refer to manifestations of alienation online. Stemming from digital labour studies (Krüger and Johanssen 2014) the focus soon bridged into social networking. A good example of this bridging can be seen in Fuchs and Sandoval's *Framework for Critically Theorising and Analysing Digital Labour* (Fuchs and Sandoval 2014). Initially exploring the dimensions of paid digital labour, the authors extend the analysis into the realm of unpaid labour within content production in social networks. P.J. Rey's paper *Alienation, Exploitation and Social Media* (Rey 2012) explores the mechanisms by which capitalism has come to exploit social media. Rey's task involves demonstrating how alienation exists within social networking as a dynamic of value extraction. This approach is also used by Christian Fuchs (Fuchs 2013; Fuchs and Sandoval 2014) in most of his work. By contrast, Krüger and Johanssen's *Alienation and Digital Labour—A Depth-Hermeneutic Inquiry* (Krüger and Johanssen 2014) examines alienation through a survey of prosumer's comments about social network's themselves. Here alienation is demonstrated through the effects of the social network system's activities, rather than through the dynamics of labour and surplus value extraction. Of course, if alienation can derive from unpaid digital labour, the possibility arises that alienation can be found wherever unpaid digital labour occurs. Here we find Marc Andrejevic's *Surveillance and Alienation in the Online Economy* (Andrejevic 2011b), which extends the analysis of alienation beyond social networking into general online activity. This paper shows a third approach to explaining digital alienation by focusing on exploitation, in contrast to the previously mentioned papers, which focus on value extraction and coercion. What all these analyses demonstrate is that the nature of alienation online necessarily diverges from the account of alienation in earlier, pre-digital, analyses. These divergences reflect the differences in structures of production and value-extraction between analogue and digital socio-technical systems. These differences are significant to the degree we may warrantably talk of a distinct "digital" form of alienation.

In Marx, alienation is the result of labour activity coerced into alienated forms in order to produce products estranged from the producer (Ollman 2001). The political dynamic is the extraction of value from controlled and structured worker activity. Historically, analysis of digital alienation has focused on accounting for the traditional mechanisms underpinning alienation within a digital context. There has been an unspoken consensus that an account of digital alienation requires identifying the same structures and mechanisms within the digital context as Marx identified within the factory. Here the concern is to understand digital alienation by analysing it as the result of conditions considered necessary for alienation - coercion, labour and estrangement from product. With regard to coercion, the difficulty is whether people who freely choose to use social networks like Facebook can be described as coerced. The concern with labour is whether people's unpaid production of content in social networks can be described as labour. Finally, if people seem to be expressing themselves within social networks, the question arises as to how can they estranged from the output of their activity.

At one extreme researchers, such as P.J. Rey, have argued that the differences between the Victorian factory of Marx's analysis and modern digital activity are so great that alienation is a questionable concept within a digital context (Rey 2012). Rey argues that the products of digital labour in social networks are not alienated because creation of this content is freely chosen and creative. Referring back to Marx's categorisation of imagination as a distinguishing characteristic between animals and humans, Rey suggests that the creative nature of social content production renders the output unalienated. His view is that the creativity involved in social network content creation allows the producer to recognise themselves within their output. In addition, the free choice to engage in social networking means this labour is uncoerced. Rey does accept there is some degree of exploitation involved because social networks derive financial value from this output without financially compensating those who produced it. However, he argues this exploitation is mild because producers do receive compensation in other forms of capital. Rey argues that social network users are compensated because they retain use of their output for their own purposes. They can therefore use the content they produce to generate social and cultural capital. His position is that the non-economic value derived is so great that any exploitation is "relatively minimal" (Rey 2012, 415). Furthermore, any exploitation present is, Rey argues, further diminished by the unalienated nature of prosumer output. Rey acknowledges that social networks also derive value from surveillance of user activity, usually without users being aware of it. However, while he sees this as mildly exploitative, he does not consider it alienating. Rey's position is that digital capitalism can maintain the inequalities and power structures within society identified by Marx, but without the need for alienation, or even very much exploitation.

It is notable that, while recognising that social networks extract value from user surveillance, Rey does not extend this recognition to the fact, noted by others, that such surveillance is almost universal throughout the internet (Turow 2011; Mayer-Schönberger 2009). A 2012 study of the world's busiest websites revealed that 94% themselves engaged in some form of user surveillance, half of whom also allowed unidentified third parties to engage in such tracking through their sites. The same study also found that 91% of these sites changed their content to match their understanding of the user (Sandoval 2012), something impossible without a pre-existing knowledge of that user; knowledge which can only have come from previous surveillance. User activity in other parts of the internet, such as search and reading, does not generate cultural or social capital, but is still subject to the same levels of commercial surveillance. Following the logic of Rey's analysis, this renders such surveillance much more exploitative. In general, Rey's analysis treats technology as invisible and as permitting users to fully express themselves in an unmediated fashion. While Rey recognises that surveillance occurs, he fails to take into account that much it is used to tune and filter the online environment surrounding the user. Users are presented with "personalised" choices, links and content based on the results of covert surveillance as much as on the content they produce; something often referred to as the *filter bubble* (Nagulendra and Vassileva 2014; Resnick et al. 2013). People are therefore not able to make free choices or even fully express themselves, because the technology available to them is not value-neutral, but tuned to commodification (Koh 2011; Erlandsson, Boldt, and Johnson 2012).

Andrejevic's analysis of digital alienation is founded on just this consideration. All internet users are subject to pervasive universal surveillance by commercial enterprises (Omar Tene and Polonetsky 2012; Turow 2011; Sandoval 2012; Deighton and Kornfield 2012). The value of this surveillance far exceeds that derived from social network content creation (Deighton and Kornfield 2012; Google Inc. 2014; Facebook Inc. 2014; Turow 2011; Mayer-Schönberger 2009; Omar Tene and Polonetsky 2012). Initially this information was used only to tune advertising delivery (Deighton and Kornfield 2012; Turow 2011). However, this information is now also used to tune the delivery of news on many sites (Turow 2011) and for political manipulation (Abse 2012). Users have no choice over whether their activity online is recorded, processed and used, nor do they know who by (Sandoval 2012). This constitutes, for Andrejevic, alienation. His argument is that the lack of choice over whether to be surveilled or not constitutes a structurally-embedded coercion. He further argues that the lack of knowledge about this surveillance constitutes an epistemological alienation. Finally he argues that the use of this information to alter content in an effort to manipulate the user fits Marx's definition of alienation as an estranged power structure working against the individual (Andrejevic 2011b, 2012).

In contrast to Rey's position that exploitation is mild because the user derives non-economic use value from the content they create, Christian Fuchs has argued that exploitation is either present or not, and cannot be present in variable degrees (Fuchs 2013). One cannot be a little bit exploited. Fuch's work tends to focus on the mechanisms of value-extraction within a digital context. Fuch's position is that any activity conducted by someone which can be used to generate economic value is labour. He seeks to bring together the competing positions held by Andrejevic and Rey by arguing that Rey is focused on subjective feelings of alienation whereas Andrejevic is focused on the objective conditions of non-control and non-ownership. However, Fuchs firmly comes down on the side of alienation being objectively present, arguing that the purported social use-value that content creators derive from their work hides the true commodity character of social networking (Fuchs 2014). He identifies two dimensions of value within social networks - the value of created content and the value of user presence. Here Fuchs agrees with Andrejevic that the users of social networks are themselves treated as commodified products which are then sold.

Both Fuchs and Andrejevic limit their conception of the use of personal data to the realm of advertising delivery. While the first use of this information was indeed to tune content, especially advertisements, to the user profile, this information is now also sold for other purposes, including political manipulation (Abse 2012), credit scoring (Schmitz 2015; Mierzwinski and Chester 2013), housing and employment (Citron and Pasquale 2014) and news delivery (Turow 2011). It is worth noting that both Facebook and the international trade body for online advertising, the Internet Advertising Bureau, agree with this assessment of where the real value lies in commercial online surveillance (Facebook Inc. 2014; Deighton and Kornfield 2012). In comparison with this vast and pervasive surveillance industry, user-generated content within social networks is a trivial consideration. Under this analysis, alienation is a pervasive and unavoidable adjunct to almost all digital activity.

Rey, Andrejevic and Fuchs all approach alienation within a digital environment by focusing on Marx's mechanisms for its production and explaining how and where these mechanisms can be found online. While general commercial surveillance is mentioned, it is not really the central focus of their analysis, nor does it alter their approach. Our position is that we can better account for digital alienation if we can liberate ourselves from the form of Marx's account. Marx provided an analysis of how alienation occurred within a particular historical and technological context. As we have seen from the above, we encounter problems if we assume that this is the only mechanism by which alienation can occur or that all of these traditional mechanisms are necessary. The argument herein is that the features of digital alienation are so different from traditional alienation that a new account is necessary.

5.2.3. How alienation occurs online

In defining alienation, Marx considered two factors, the nature of alienation and the means by which it is produced. The nature of alienation is that the individual is disconnected from the products of their labour by property ownership rights; they are alienated from ownership of both the product and the means of producing it. This constitutes the material base of alienation and is the product of power relations governing the production process. Marx's account involved material coercion by controlling access to the means of survival so as to force people into alienating labour. Analysis of exploitation on the internet has been distracted by the apparent lack of coercion motivating online activity and by the appearance of self-expression in social networks. However, our analysis becomes less complicated if treat social networking within the broader context of pervasive digital surveillance. Here we recognise that, while content production in social networks is voluntary and can be self-expressive, it is just one type of action within the wider class of voluntary and self-serving digital activity which includes search, shopping, email, use of maps, health trackers, life loggers and other digital services, not to mention general web surfing. This is important because the range of digital activities will continue to spread until it permeates most of our environment (Rader et al. 2010). Because of this it is essential to treat the current state of affairs as an intermediate process moving towards more ubiquitous computing. Our analysis must recognise that the political and economic structures which affect us within the current digital domain are on a trajectory to dominate our entire existence, offline as well as online. It is important, therefore, to recognise that the frame of analysis cannot limit itself to voluntary activity knowingly making use of digital services. The infrastructure being created now will one day support smart cities, the internet of things, and digital devices implanted within our bodies. Our entire existence will become mediated through digital services within a few decades (Rader et al. 2010).

Thus the place of labour as seen in a traditional account of alienation becomes problematic when value is being extracted from broad-spectrum use of digital services for life in general. Assuming that labour is a necessary precondition for alienation requires explaining how all activity using digital services constitutes labour despite the fact it generates no obvious income and may not even be anything more than a traditional activity, like walking or driving, which has been supplemented with a digital component. Certainly the argument of remuneration in the form of social or cultural capital is inapplicable with reference to activities which do not involve any form of communication, such as using search engines or passively reading a website, yet value is extracted from these activities by others via commercial surveillance (Andrejevic 2012; Brooks 2005; Deighton and Kornfield 2012; Rubinstein 2013; Schneier 2015; Turow 2011; Palmås 2011). If we redefine 'labour' as referring to any activity from which value may be drawn by any party, as Fuchs does (Fuchs 2013, 2014), then almost all activity becomes labour and the term ceases to provide any real distinction from other mode of activity. We think it is better to abandon the issue of whether online activity is labour or not. There is nothing within Marx's description of alienation which requires that it must, of necessity, derive from labour. 'Alienation' in Marx is not a single concept, but a translation of two terms, *Entfremdung* and *Entäusserung*, which can also be translated as 'estrangement' and 'externalization' respectively (Ollman 2001). These terms are applied to a variety of phenomena, including internal mental states, property relations and societal structures. It is true that Marx attempts to provide a systematic analysis of political economy based on the concept of alienated labour in his early work, but that attempt is incomplete (Wood 2006). In his later works, alienation becomes a descriptive term which is applied to multiple phenomena. There is nothing in his usage which locks alienation to labour except as a historically contingent feature of nineteenth century capitalism (Wood 2006). All that is required by Marx's account is that there be human activity and that this occur within certain types of unequal power structure within the field of economic competition.

On this basis, we propose to focus on digital alienation as a product of property relations regarding data. Surveillance is a process of data acquisition; some generated as the output of online surveillance monitoring systems and some data taken from elsewhere, such as the passenger name records used for international travel, geo-location data and credit scores (Abse 2012). The common element all these data elements have is that they are held to pertain to the same individual[12]. The dataset created is termed a *personal profile*, as opposed to *group profiles* (Hildebrandt 2008; Kennedy 2008). The personal profile is a digital representation of an individual. It is the central commodity of the surveillance economy. Each organisation which holds a personal profile subjects it to algorithmic analysis and manipulation in order to extract value from it. The term used in the industry is to "monetize" it. It is this profile which is used to tune content and for purposes of manipulation. All actions using personal data draw that data from the personal profile. Such use constitutes Marx's concept of an environment which reflects back on the producer estranged output (Andrejevic 2012). In that surveillance technology produces the personal profile as a commodity, it is a type of production process. The raw material for this production process is the activity of individuals (van der Hof and Prins 2008), which is used to produce personal profiles. This production

[12] This belief may be mistaken, it is not always possible to distinguish between a person and a device; and cases of mistaken identity also occur.

process is not owned by those who generate the activity which feeds it. This is the basis for alienation from ownership of the means of production. The surveillance process is hidden and unwelcome (Culnan 1993; van der Hof and Prins 2008) and therefore represents an unequal, coercive and exploitative power structure (Andrejevic 2012). We may view the personal profile as a field of contention between commercial surveillance companies and those who use their products on one hand opposed by individuals and privacy advocates on the other.

The essential starting point to all forms of alienation is individual activity. We therefore believe digital alienation may be best understood by examining the mechanisms by which an individual's digital activity is alienated. Here we must focus on the nature of personal action within a digital context, the mechanisms by which the personal profile is generated, and the use to which it is put. As mentioned above, the first task is to dispense with the need for a concept of labour. In Marx's analysis labour was the term used to distinguish activity which supported alienation from activity which did not. Labour supported alienation because it was activity which occurred within, and was shaped by, exploitative power structures. However, no such distinction between labour and non-labour exists online because all activity is surveilled and exploited (Schneier 2015; Turow 2011). Not only does discarding the need for labour ease our analysis, it helps to direct our attention to the ubiquity of digital surveillance. Instead we will define human activity within a digital context in terms of people's intentions and expectations. To do this we will distinguish the two targets of surveillance; communicative activity and everything else. We will refer to these as the *digital persona* and the *data shadow*, respectively.

'Digital persona' is our term for the body of digital material created by an individual in acts of online communication. The digital persona includes blogs, comments, product reviews, tweets and other social network postings, together with any other conscious communication by an individual within a digital context. Thus the digital persona is created by the individual to express and communicate. The digital persona is not a direct or unmediated reflection of the personality, but a creation through which the individual seeks to represent of an aspect of themselves. The disconnect between the offline and digital world permits people to exaggerate or repress particular aspects of their personality (Suler 2004). For example, introverts may use the digital persona to compensate for difficulties they have in face-to-face interactions (Amichai-Hamburger, Wainapel, and Fox 2002) while extroverts often use it to confirm pre-existing characteristics (Wang 2013). In other cases, people develop new personal characteristics online so that they can incorporate them into their offline personality (Mehdizadeh 2010). In all cases, what is revealed or portrayed is further influenced by previous experiences online, especially concerns over privacy and security (Yao and Linz 2008; Krasnova et al. 2009). We derive the term 'persona' from C. G. Jung's concept of the persona as a creation of the ego designed to represent a subset of that ego within specific social circumstances (Jung 1977). The same idea is used within a sociological perspective in Goffman's *The Presentation of Self in Everyday Life* under the term 'masks' (Goffman 1990), which outlines his Dramaturgical Theory, a sub-set of Symbolic Interactionism (Ritzer 2011).

The term 'data shadow' was first used by Alan Westin in *Privacy and Freedom* (Westin 1970), but has entered into general use both in computing and privacy discussions. It refers to the information generated by someone as a side-effect of their use of digital technology. These days this includes log files, access records, search histories, movements between and within web sites, mobile phone location records and all financial activities not involving cash (Koh 2011; Hildebrandt 2008; Brooks 2005). Thus the term 'data shadow' refers to all digital information pertaining to an individual which they did not consciously and intentionally create for communicative purposes. This information may have been generated by the user for other purposes, such as their *click-stream history*, which is a record of their mouse click activity within a website (van der Hof and Prins 2008), or their *search history*, a record of all the searches they have made in a given search engine. Elements of the data shadow can also be generated through the monitoring and recording of user activity by other systems. For example, web server log files, containing records of every file request, constitute data generated by the system about the user. The term 'data shadow' includes the material used to commodify users within social networks, but also applies outside social networking. Data shadows may be created through any and all use of digital technology.

Data shadows are created by a network of commercial surveillance agencies whose tracking technologies permeate digital services (Mayer-Schönberger 2009; Turow 2011; Deighton and Kornfield 2012; Brooks 2005; Lazarus 2015). Very few of these agencies are known to the public (Brooks 2005; Schneier 2015). Some, like Google and Facebook, are well known because of their public profile as digital service providers, though their activity as commercial surveillance agents is less well known, even though it drives their profits (Google Inc. 2014; Facebook Inc. 2014). Others, such as

DoubleClick, Acxiom, Experian and BlueKai, are known to industry analysts and privacy advocates as a result of their scale and reach. However, the majority, such as ClickTale, Optimzly, Kiss Metrics, Info Group, Ace Metrics, Crazy Egg, Site Meter, Moz, Adgistics, People Metrics, Data Dog, Data Mentors, Extrawatch, Inspectlet, eDataSource, Prognoz, and literally hundreds of others, are unknown outside the specialist profiling industry. No one knows how many of these agencies there are, or what they do, but it is known they combine the data they gather with information from other sources to create detailed profiles on literally hundreds of millions, if not billions, of people (Brooks 2005; Federal Trade Commission 2014; Wayne 2012). The commercial surveillance industry is much larger in terms of economic value and user-base than any other online industry (Deighton and Kornfield 2012; Schneier 2015; Turow 2011).

This universal commercial surveillance means there is no way to use most digital services without being surveilled (Turow 2011; Wayne 2012; Schneier 2015; Federal Trade Commission 2014; van der Hof and Prins 2008). For most digital services there is no alternative provider who does not practice surveillance (or permit others to do so) within the service stream (Turow 2011; Stole 2014). However, lack of choice most strongly stems from lack of knowledge. We are simply unaware of when we are being surveilled, who by and for what purpose (van der Hof and Prins 2008; Omar Tene and Polonetsky 2012). Obviously, one cannot exercise choice over things one is unaware of. As we have seen, this lack of choice has been held to constitute coercion by Andrejevic and Fuchs, but not by Rey. Lack of choice as coercion has a long history of support in philosophy. For example, Aquinas argues that coercion occurs when actions by one person mean someone cannot act otherwise (Aquinas 1920). However, this position was challenged in the twentieth century by the position that coercion requires communication between the coercer and their target, usually in the form of conditional threats (Airaksinen 1988). Under this view coercion is a communicative act, not a contextualising situation. This is the position currently supported in much legal practice, especially in the USA (S. Anderson 2010). However, since the 1980's arguments have re-emerged in support of structural coercion; the creation of situations in which one is prevented from selecting alternative courses of action (Rijt 2012). Here the focus is shifted to the coercer's intentions to remove choice from another (S. Anderson 2008). This accords with much of Marx's analysis, in which he focuses on the general circumstances of capitalist society as coercive in the sense of removing freedom (Wood 2006). Clearly, hiding surveillance so that people cannot avoid it constitutes removal of choice and diminution of freedom. Thus it is possible to argue from this perspective that the lack of choice to avoid surveillance constitutes coercion. However, we must recognise that this position is not in accord with how many, especially in jurisprudence, understand the term.

Lack of choice, even coercion, does not automatically mean that the output of a productive process is alienated. We wish therefore to explore the mechanism by which digital activity becomes alienated. Since we have two forms of digital data, the digital persona and the data shadow, two accounts are necessary. We shall commence with the alienation of the digital persona.

5.2.4. Ego, Affordances and the Digital Persona

People use Web 2.0 technologies to create their digital persona. The process by which they do this and the persona they create are alienated. We therefore need an account of the mechanism by which people do this and how alienation occurs. Central to our account of how the digital persona is alienated is the view of technology as a socio-technical system (Ikonen et al. 2010). A technology may be composed of multiple artefacts and may be "read" or understood in different ways (Hutchby 2001; Heidegger 1977). The nature of the "reading" depends on the person, their social environment, past experiences and other factors, all of which are constrained by the functional capabilities of the artefacts in question (Norman 2002). We therefore need a conceptual framework which holds all the dynamics which are at play in a person's understanding and use of a technical system. We will use the concept of *affordance* to explain the interaction between people and the technical artefacts.

The concept of affordances originates with James Gibson's account of how animals perceive and understand their environment in *The Ecological Approach to Visual Perception* (Gibson 1986)

> "The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill… It implies the complementarity of the animal and the environment… [affordances] have to be measured relative to the animal. They are unique for that animal. They are not just abstract physical properties." (Gibson 1986, 127)

Gibson was arguing against a reductionist understanding of perception and for a perceptive process within all animals in which perception itself is not merely a process of physical activity onto which understanding is overlaid *post hoc.* Instead, he argued that the perceptive process itself incorporates cognitive elements such as motivation, environmental context and past experience into the act of seeing.

The concept was applied to ICT analysis by Ian Hutchby in *Technologies, Texts and Affordances* (Hutchby 2001), in which he describes technologies as

> "texts which are written in certain ways by their developers, producers and marketers, and have to be read by their users or consumers. The writers of these technology texts may seek to impose particular meanings on the artefact, and to constrain the range of possible interpretations open to users. Users, by contrast, may seek to produce readings of the technology texts which best suit the purposes they have in mind for the artefact… Neither the writing nor reading of technology texts is determinate: both are open, negotiated processes. Although there may be ways that technology texts have preferred readings built into them, it is always open to the user to find a way around this attempt at interpretive closure." (Hutchby 2001, 445)

We may thus see affordances as a field of competition in which the owners of a technology compete with the users of that technology for domination of the affordances which dictate how that technology is understood and used. Donald Norman explores this competition over technological affordances in *The Design of Everyday Things* (Norman 2002). In Norman's account, we use affordances to build conceptual models of how things work. Any technology involves the interaction of two conceptual models; a design model and a user model. The design model is the conceptual model the designers held when they built the technology, and in accord with which they try to construct the artefact. The user's model is the conceptual model users have of that same technology. Norman is concerned with what happens when the two models clash or diverge. According to Norman, there is no necessary convergence between the user's mental model and the designer's. In fact, in Norman's view, the two model's clash most of the time. Using Norman's framework, we suggest that the user model conceptualises the Web 2.0 services people use to express their digital personas as private, unmediated and natural. The user model fails to recognise the degree of surveillance and the degree to which their activities are mediated through a technology designed for data gathering and commodification. Users also fail to recognise the degree to which surveillance is used to filter and control the content they see in social networking and news sites and in advertising. Instead, users see the content presented to them within social networks as somehow neutral, unmediated and unsurveilled (dos Santos Brito et al. 2013). In contrast, service providers, such as Google and Facebook, show evidence of believing that users have the same conceptual model as designers. They have countered concerns over online privacy by stating users have no expectation of privacy and accept that the material they create will be processed for purposes of commodification (Fogel 2014).

There are numerous studies which demonstrate that users manipulate their self-expression online in order to convey specific characteristics and control the image others have of them (Krämer and Winter 2008; Mehdizadeh 2010; Wang 2013; Martin and Nicovich 2013). In our terminology we may say people use Web 2.0 technologies to construct their digital personas. Their understanding of what can be expressed, the values determining what should be expressed and how this is to be done are determined by the affordances users perceive in these technologies (Sherman 2001; Zhao, Grasmuck, and Martin 2008). These affordances constitute what Goffman describes as the "props and tasks" (Goffman 1990, 143) which dictate what persona[13] is appropriate and the "expressive resources" (H. Miller and Arnold 2001, 74) available from which to construct it.

Unfortunately for users, Facebook and similar Web 2.0 systems are not designed for people to portray themselves in any manner they may choose. Instead, Facebook and similar systems divide personal characteristics into a set of discrete data points, such as preferred objects of consumption, marketable skills, and approvable attitudes (Hull, Lipford, and Latulipe 2011). Furthermore, qualitative characteristics, such as friendship, are reduced to quantitative values, such as the number of likes or followers. Facebook's affordances, in particular, suggest to users that their digital persona is a true reflection of their identity, yet is at the same time something to be constructed, managed and enhanced (Gershon 2011). Facebook openly expresses the neo-liberal concept of a "personal brand," in which a person creates a commodified public image as the repository of their social capital (Lair 2005). The affordances of Facebook present the individual as composed of consumption patterns (such as preferred movies, books and music) and patterns of association (as shown through one's likes, friends and photos). These are dimensions of analysis more suited to processing

[13] Goffman's term is 'performance' (Goffman 1990).

for advertising than developing an understanding of the whole person. There is good empirical evidence that this model conflicts with the affordances the user brings to Facebook. In many cases users seek to express themselves in ways restricted by the affordances Facebook imposes, resulting in dissatisfaction, resistance and disuse (Gershon 2011; Sherman 2001; Martin and Nicovich 2013; Krasnova et al. 2009).

In using affordances tuned to atomising, quantifying and commodifying the depiction of people, social media systems like Facebook alienate the digital persona. Rather than a free expression of the self, users are forced to display only those characteristics which are commodifiable. These characteristics are then embedded in a manipulated content environment which reinforces and promotes ongoing commodification, and therefore embeds the alienated digital persona within an alienated social environment.

5.2.5. Alienation and the Data Shadow

The mechanisms by which the data shadow is alienated are straightforward compared to the digital persona and commence with unavoidable, hidden, ubiquitous commercial surveillance (Schneier 2015; Turow 2011; Sandoval 2012). In that the digital world is permeated with unknown entities gathering unknown information to use for unknown purposes (Federal Trade Commission 2014; Wayne 2012), the digital environment is self-evidently epistemologically alienated from the user. The material base of the commercial surveillance system supports a superstructure devoted to exerting power over the individual by influencing their behaviour directly, or by influencing decisions made about them by other people (Brooks 2005; Schmitz 2015; Omer Tene and Polonetsky 2013). This is achieved through personalization of content (Nagulendra and Vassileva 2014), such as the advertising (Sweeney 2013) and news (Turow 2011) to which people are exposed.

The unavoidability of commercial surveillance is made possible by the lack of ownership or control users have over digital services. It is known that, in general, people do not like commercial surveillance or content personalization (van Doorn and Hoekstra 2013; Culnan 1993). Commercial surveillance therefore constitutes the exercise of power over individuals and a diminution of their freedom, another manifestation of alienation (Andrejevic 2011a). Furthermore, the knowledge that unknown surveillance is occurring, in combination with lack of knowledge about how that information is used, has a chilling effect on people's online activities (Solove 2004; van der Hof and Prins

2008; Marthews and Tucker 2014). In effect, people are alienated from their own actions online before they perform them. In that this chilling effect also applies to how people communicate online, ubiquitous commercial surveillance further alienates people from each other.

5.2.6. Digital Alienation – the complete picture

We are now in a position to provide an account of how the four dimensions of alienation occur. First, users are alienated from their productive activity through restricted affordances within expressive Web 2.0 technologies which promote a commodity fetishism of personal characteristics and interpersonal relationships. This is made possible by an alienated power structure which is designed around treating users as commodities (Andrejevic 2012; Fuchs 2013). Users are alienated from non-expressive activity by the presence of ubiquitous hidden surveillance systems. Thus users are alienated from all forms of digital activity. Second, users are alienated from the products of their digital activity by property relations. These grant service owners the right to reuse user-produced content for their own purposes and to process both the digital persona and the data shadow in order to construct personal profiles. Users are further alienated from the products of their own activity since the personal profile is used against them, either to manipulate their behaviour or to influence how others treat them. In addition, the abandonment of the open standards which created the web means that the products of user activity are imprisoned within data silos owned by service providers (Dyhouse 2010). Thus, you may close your Facebook account, but you can't move it to another social network. Third, users are alienated from each other by the necessary mediation of fetishizing social networks and by the chilling effect of ubiquitous surveillance. Finally, users are alienated from themselves and their own human potential in three ways; through the imposition of fetishizing affordances promoting the concept of the personal brand, through their limited control over their own digital persona, and through the use of personalization technologies which confine the user's ability to discover the unexpected, the unusual, and the uncommodified.

5.2.7. Solutions

No solution exists today which can resist these patterns and structures of digital alienation. However, a number of technologies exist which can form part of a solution, while the design principles to complete the solution are understood. Two related characteristics support the existence of digital alienation, lack of choice and lack of power. The solution is therefore to provide restore choice and empower the user. In our view solutions that look to regulation, such as data protection and privacy laws, merely perpetuate a hierarchical structure which keeps people in a powerless position. Instead of companies deciding what to surveil, we merely pass the decision to legislators. Given the history of government digital surveillance (Angwin 2014) there is nothing to suggest this improves matters. In addition, the impossibility of a single legislative framework for the entire internet (Zekos 2006; Wilske and Schiller 1997; Reidenberg 2005; Dainow 2013) means surveillance companies can simply move to more conducive regimes. Furthermore, centralised storage of personal data is frequently subject to leaks (Liu et al. 2011; Madejski, Johnson, and Bellovin 2012; Koops and Leenes 2014), so we are opposed to centralised storage of any fashion, never mind under what rules.

The first task in combating alienation must be to remove coercion from the situation by giving users the choice over whether to be surveilled and for what purpose. A number of technologies exist which can offer elements of this solution. Anonymizing systems such as such as TOR (McCoy et al. 2008) and TextSecure (Oppliger 2014) enable users to avoid being tracked while using the existing internet. These need to be extended and built into a comprehensive set of easy-to-use systems which can wrap browsers and other applications in a protective and intelligent layer which negotiates and controls what data is accessed by what services. Protocols like the W3C's Platform for Privacy Preferences (PPP) (Cranor et al. 2002) can form the basis for such communications. Design of data gathering systems should follow principles of privacy preservation, such as those developed by Marc Langheinrich (Langheinrich 2001), one of the authors of PPP. Such technology would enable users to control how much information is gathered about them and thus how much personalization is possible. This de-alienates the productive technology by putting control in the hands of the user and de-alienates their digital environment by permitting them to control or prevent personalization.

While these solutions restore choice to the user, they only partially redress the balance in an existing system which is structurally inequitable. The long-term solution must therefore be to move personal data storage, and therefore ownership, into the hands of the users. Here the solution is to reverse the cloud architecture. Currently, centralised systems run analyses of locally held data. We would reverse this, so that personal data is held by the person in their own devices. Effectively each person, or home, would operate their own data store. Following Langheinrich's privacy-preserving design principles (Langheinrich 2001), devices would, wherever feasible, store their own data. A personal server or gateway would provide the interface between digital service providers and the user's personal data. This gateway would be able to negotiate access for services and prepare personal data for access. This pre-processing would anonymise the data to the degree selected by the user for that type of service. We envision this system working in a manner similar to hierarchical protection domains (or "security rings") within chipsets. These create a series of layers within which particular software operations can be confined so as to shield the system from inappropriate operations (Karger and Herbert 1984). Corporate digital services could still be centrally managed and owned, but their computations would have to call on the individual's own data store rather than house it on corporate servers.

This is, however, merely a collection of artefacts. As a socially-embedded system, technology needs more than just hardware if it is to be adopted. The additional component required is therefore societal structures promoting and maintaining such a system. A network of local technicians would be required to maintain and develop such systems, provide advice and training, lobby regulators for support and so forth. Here we think the basis lies in recognising the value of personal data. For example, the value of a Facebook user is between $US40 and $US300 (OECD 2013b). If personal data has value for service providers, let them pay for it. A system of micropayments for access to personal data would create a data economy enabling individuals to earn money through the gathering and storing of their own data. Support agencies, such as technical staff and software vendors, can then be remunerated through a share of this income. Such a system would permit the development of an intermediate layer of data vendors who can store and provide personal data on the user's behalf, according to guidelines provided by those users, or remotely maintain data held in the home. Such a system permits of multiple organisational models. Community groups could operate such services. For example, people who share the same set of data access protocols could form cooperatives to manage storage and access to their member's data. As yet, such technology does not exist. However, the hardware is already in place. Personal cloud storage devices have been available for several years. These permit users to store their data in their home while still being able to access it remotely. The missing components are therefore the micropayment and data negotiation systems. Protocols exist which can handle both, they merely need to be implemented as working products.

We need to bear in mind that the digital service infrastructure we see today is merely a step towards a digital environment of ubiquitous devices; embedded within our bodies, throughout our homes, offices, cars and public spaces. A critical evaluation of current data practices must consider this long-term future and seek emancipatory paths within it. As we have seen, digital alienation is the product primarily of inequitable power structures which intentionally deny users control, or even knowledge, of what is being done to them. The motive power of these structures is the economic value of personal data. If digital services are to align with individual needs, we cannot avoid personal data being processed. The solution is therefore to develop systems which pass some of that value back to the user. Doing so gives the user power and makes them a viable partner for other organisations who can earn a living by controlling access to personal data on behalf of the user. Giving the individual control over their personal data emancipates them from subjection to hegemonic digital capitalism by permitting them to negotiate the terms of the relationship they have with their digital service providers.

- end of published paper –

5.3. Next Steps

Having developed a social account of ICT's, the next step in the research project was to use this to consider the way each of ETICA's eleven technology groups could restrict human autonomy. However, as detailed in the next chapter, autonomy is a complex and contested concept. We identified six different classes of account regarding what constitutes autonomy in the philosophical literature, but no recognition of the variability of the concept in legal or political discourse. In general, the philosophical approach is concerned with the internal (mental) states one must enact in order to be autonomous, while legal and political accounts are focused on external aspects, such as lifestyle, intersubjectivity and rights, without reference to internal activity. The result in practical terms is that one may legitimately claim to be living autonomously while living in a manner another would not consider autonomous. Indeed, some definitions of autonomy, such as coherentist accounts, would regard this as inevitable and appropriate, as would most legal and political accounts. It is not, therefore, appropriate for our research to select a particular account of autonomy as the only correct one and restrict analysis to only that form. Indeed, we consider it to be one of our most significant contributions that we introduce into ICT ethics and computer science an understanding that different, yet legitimate, accounts of autonomy exist and are impacted in different ways by different species of ICT. This was published in 2017 as "Threats to Autonomy from Emerging ICTs" in *Australasian Journal of Information Systems (21)*.

6. Autonomy under Emerging ICT's

The paper contained in this chapter answered the main research question:

> "What are the possible threats to human autonomy generated by emerging ICT's?"

The paper also answers the following secondary research questions:

> "What are the emerging ICT's?"
>
> "What is human autonomy?"

Emerging ICT's cover a wide range of technologies, yet human autonomy is potentially threatened by all of them in some way. However, as indicated above, because there is no single agreed definition of autonomy, this paper considered the ways in which different accounts of what constitutes human autonomy are impacted by each of ETICA's ICT groups. From this range of threats we derived properties which any ICT must exhibit in order to threaten human autonomy. Finally, the paper demonstrates how the range of definitions of autonomy creates problems for customary approaches to value-sensitive design, indicating a need for greater flexibility when attempting to improve the ethical status of emerging ICT's.

The rest of this chapter contains the paper as published. Minor modifications have been made to ensure the writing style matches that of this thesis (such as changing "I conclude" to "we conclude"). The citation style has been changed to maintain consistency with the rest of the thesis. In keeping with the need to present the published material in the format of a coherent sequence of chapters, journal meta-information, such as keywords and categories, have not been reproduced here. The original publication did not number headings. Numbering has been added to maintain consistency with the rest of the thesis.

6.1. Threats to Autonomy from Emerging ICTs.

Australasian Journal of Information Systems 21, no. 0 (2017). https://doi.org/10.3127/ajis.v21i0.1438

6.1.1. Introduction

This paper examines ways in which human autonomy can be threatened by emerging ICT's. Emerging ICT's embrace a wide range of technologies, including autonomous systems, intelligent environments, bio-electronic implants, robotics, and artificial intelligence. We will explore the different ways in which each ICT threatens autonomy to show that the nature of the threat, if any, depends on the version of autonomy being used. From this survey of threats we will derive some properties which any ICT must exhibit if it threatens human autonomy, no matter how it is defined. Finally, we will argue the range of definitions of autonomy demonstrates the need for greater flexibility when attempting to improve the ethical status of emerging ICT's.

In order to discuss the complete range of emerging ICT's it is necessary to have a system by which to conceptualise and organise them. We shall use for these purposes the output of the ETICA project (Stahl 2011b), which attempted an examination of the complete range of emerging ICT's, developed an ordered taxonomy (Ikonen et al. 2010), and researched the ethical concerns associated with each (Heersmink, van den Hoven, and Timmermans 2010).

6.1.2. Methodologies for evaluation of emergent technology

To count as "emerging" (as opposed to "possible") an ICT must be sufficiently developed that we can understand its general features, but which has not yet achieved maturity in all aspects. Since the 1960's the predominant model for such analysis has thus been to treat technologies as "socio-technical systems" (Trist 1981; Lamb and Kling 2003). This position contrasts with the other major treatment of technology, technological determinism, which treats technology as affecting society, but as developing independent of social, cultural, political and other "intangible" causes (M. R. Smith and Marx 1994). An emerging ICT will have little market penetration (it may still be in prototype), such that its final place and role in the market is yet to become evident; its full and final set of features are unlikely to be completely determined; the business models under which it will function are likely to be nascent; users may still be working out patterns of usage; and the regulatory framework is most probably undeveloped.

Autonomous cars provide a simple example of an emerging ICT which illustrate these uncertainties. Google's autonomous cars are still in prototype (Solveforx 2016), while Tesla vehicles are just beginning to penetrate the market (Statistica 2016). In both cases, the systems are still in development such that the final set of functions is indeterminate (Statistica 2016; Musk 2016). The regulatory framework for autonomous vehicles has seen some development in a few places, such as California, but most countries haven't even begun to think about the technology. In the USA, for example, as of March 2017, only eleven states had any regulations regarding this technology, while the majority of attempts to introduce legislation in the USA since 2012 have failed

(National Conference of State Legislatures 2017). Business models for the sale and use of this technology are still undetermined. For example, Tesla recently proposed models in which people would be able to send their Tesla car out as an automated taxi when they're not using it (Musk 2016). In the meantime, users are still learning how to use the systems, often in conflict with how the designers intend them to be used. The most dramatic example of this disparity between designer's intent and user behaviour is the frequent driving of Tesla vehicles without human supervision, despite the Tesla's frequent statements that the user should pay attention and keep their hands on the steering wheel at all times (Consumer Reports 2016). Any attempt to understand the ethical impact of autonomous cars must therefore commence with a vision of the state of affairs when the technology has matured – with standardised features, customary patterns of usage, stable regulatory regimes and (reasonably) settled business models. Any attempt to assess the ethical implications of autonomous cars thus becomes an exercise in anticipating what the future will look like. This brings us into the remit of foresight studies.

6.1.3. Introducing foresight studies

'Foresight studies' is one of the terms used to describe the discipline of considering the future in a methodical manner. Foresight studies can be used in any discipline and so the objectives in such research vary widely (Glenn 2009b; Godet 2009; Veikko, Kanerva, and Kouri 2009). This presence within many disciplines has led to a variety of terms being used to describe the activity. Common alternative terms for foresight studies are "future studies" (most commonly used in the USA), "futures research" (Europe), "futurology" (Australasia), "prospective studies" (France), "futures field" (Europe) and "prognostics" (Russia) (Veikko, Kanerva, and Kouri 2009). A wide variety of foresight methodologies have evolved and most foresight research projects combine several methodologies.

The different methodologies used within foresight studies can be divided between quantitative and qualitative in technique and between normative and exploratory in purpose, though some methodologies bridge these divides (Veikko, Kanerva, and Kouri 2009; Haegeman et al. 2013; Gordon and Glenn 2004). Quantitative methods are used mainly in predictive forecasting and seek to produce probabilities (Veikko, Kanerva, and Kouri 2009; Haegeman et al. 2013). Quantitative methods are also used in complex modelling, for example, climate modelling. Such methods are based on the assumption that future development is a continuation of past trends. While they provide a continuity from past to future, they cannot allow for disruptive technologies or unexpected developments. Some argue for the need to support quantitative methods with qualitative ones (Veikko, Kanerva, and Kouri 2009; Haegeman et al. 2013) as they are better suited for the anticipation of abrupt or unexpected changes (Ogilvy 2002; Veikko, Kanerva, and Kouri 2009). Foresight research also distinguishes between methodologies on the basis of whether they are normative or exploratory. Normative forecasting is concerned with whether a particular future is desirable and the ethical status of decisions which would lead to it. By contrast, exploratory forecasting explores what is possible, irrespective of its desirability (Veikko, Kanerva, and Kouri 2009). Ethical assessment of the future is rarely a formal component of foresight research project (W. Bell 1996; Poli 2011; Brey 2012; Clarke 2005; Grunwald 1999; Harris et al. 2010). Key foresight research figures such as Wendell Bell have attempted to determine what ethical values are universally accepted by all peoples in all cultures (W. Bell 1996). However, Bell has since been criticised for accepting as objective and universal what are merely western cultural values (Poli 2011; Dator 2011; Rubin 2011). A less controversial approach has been to incorporate ethical values as research data. Here distinct sets of exploratory and normative data are gathered. For example, the functionality of future technologies can be drawn the visions of those creating them, after which philosophical sources can be mined for ethical treatments of these functions. This was the approach taken by the ETICA project.

6.1.4. ETICA

The most comprehensive foresight research project focused on ethical assessment of emerging ICT's so far has been the EU's ETICA (Ethical Issues of Emerging ICT Applications) project, which ran from 2009 - 2011. The project's aims were to identify "significant emerging ICT's" (Stahl 2011, 1), the ethical issues these gave rise to, and to make recommendations for EU policy makers. A key output from the ETICA project was a taxonomy of emerging ICT's. Here ETICA developed formal structures for technological descriptions and scenario construction (Ikonen et al. 2010). It then used bibliometric analysis to assess published discourse regarding the ethical issues associated with these ICT's (Heersmink, van den Hoven, and Timmermans 2010). The project validated these findings with expert focus groups and surveys (Heersmink et al. 2010). Finally, ETICA developed a series of recommendations (Rainey and Goujon

2011), together with supporting philosophical and methodological expositions (Veikko, Kanerva, and Kouri 2009; Rader et al. 2010; Rainey and Goujon 2011).

ETICA identified 107 different technologies as significant and emerging, which it aggregated into eleven groups. It then identified approximately 400 ethical concerns associated with these technology groups. The aim was to collate the ethical concerns, not explore them in depth, much as an 18th century naval cartographer might map the location and outlines of a chain of islands, but not explore their interior. It is not our intention to review all of these ethical concerns, but to concentrate on issues concerning autonomy. We shall first catalogue how human autonomy can be challenged by each technology group, then elucidate the common characteristics shared by all. However, the meaning of the term 'autonomy' is much contested, so what one considers a threat to autonomy depends on what one means by the term. It is therefore necessary to briefly explore the concept of autonomy. Our aim is not to produce a definition of the term 'autonomy', but to better understand the different ways it may be used when discussing ethical threats.

6.1.5. Autonomy

Immanuel Kant was the first to apply the concept of autonomy to human beings (Schneewind 2007; Dryden 2015; Paul, Miller, and Paul 2003; Dworkin 1988). Before Kant the term 'autonomy' was a purely political one. A state was said to be "autonomous" if its laws were drafted within that state, as opposed to being drafted by a distant imperial court or some similar extra-national body (Dworkin 2015; Schneewind 2007). Kant applied the concept to the individual in *Groundwork of the Metaphysics of Morals*[14] (Kant 1785). His aim was to justify the right of each individual to make their own judgements regarding morality. For the previous 100 years, Enlightenment thinkers had been seeking arguments to combat the politically dominant understanding of ethics, in which ordinary people were seen as too weak-willed to act morally without threats of punishment and promises of reward. Under this view, society was dependant on the guidance of those few exceptional people whom God had enabled to understand and teach His moral laws. The essence of morality for everyone else was to do what they were told (Schneewind 2007). *Groundwork of the Metaphysics of Morals* set out to prove that each person drafted their own rules for their own conduct, that these rules constituted each person's moral code, and that no education was required for this ability - it was

[14] Sometimes titled "Foundations of the Metaphysics of Morals"

universally present in all humans. Kant labelled the innate capacity of all humans to determine what was morally correct as 'autonomy.' He then used this capacity for autonomy as the basis for human dignity, which then became the basis for human rights. Thus Kant's concept of autonomy created an interlinking of the concepts of individuality, freedom, morality, autonomy, dignity and politics (Schneewind 2007; Bittner 2014; Dryden 2015; Paul, Miller, and Paul 2003).

Since Kant introduced the concept, a number of differing definitions of autonomy have arisen. The range of such accounts, as well as the lack of progress towards any consensus, has led to the concept of "regimes" of autonomy (J. Anderson 2014; Dworkin 2015) – clusters of similar, but not identical, theories, such that "about the only features held constant from one author to another are that autonomy is a feature of persons and that it is a desirable quality to have." (Dworkin 2015, 8). Kant's criteria for autonomy were, by later standards, rigid and limited. Under Kant's account, people were required to determine what was right or wrong without reference to the consequences of their actions. Furthermore, under Kant, rules of conduct only counted as a moral if the person believed everyone should act the same way. There was no room for moral pluralism in Kant, or tolerance of other people acting differently (Kant 1998; Guyer 2003; R. Johnson 2014). This type of definition of autonomy has come to be known as a "substantive procedural" account. A "procedural" account of autonomy takes the position that autonomy is a specific form of thinking process (ie: follows a specific procedure). A subset of procedural accounts are substantive, adding the requirement that autonomy also involves thinking about certain things. By contrast, "content-neutral" procedural accounts try to define autonomy without reference to any particular concern or aim. Content-neutral accounts of autonomy form the majority of modern accounts (Dryden 2015) and are, perhaps, more suited to multi-cultural or pluralistic societies in that they allow individuals to determine for themselves what sources and values to consider when making ethical decisions.

The value given to autonomy varies according to the importance given to competing claims, most frequently those of paternalism and those forms of communitarianism which devalue individual choice in favour of community needs (D. Bell 2014). Kant's original concept was developed in an effort to remove external authority as a source of morality. Kant specifically identifies autonomy as oppositional to outside influence. However, we are all influenced by external forces, including family, cultural and religious influences. Consequently, outside influence and the point at which it reduces autonomy is a problem which must be dealt with by every definition of autonomy. One response has been to argue that the concept of autonomy as self-governance isolates the individual from their society and morally devalues family, community, culture and tradition (Christman 2014; Stoljar and MacKenzie 2000; Donchin 2000; Buss and Zalta 2015). These positions have given rise to relational and social conceptions of autonomy based on the necessity of human sociality.

It has also been argued autonomy, if conceived of as only certain forms of thinking, does not account for the full range of human emotional and physical experience, such as trained muscle action (e.g.: playing sport or music) or healthy reflexes, which can also express personal autonomy (Meyers 1989b, 2005). These issues collide with particular urgency where ICT's become concerned with care for the ill or aged (Agich 2003; Burmeister 2016; Dworkin 1988), but also wherever ICT's come to mediate what someone considers to be core elements of their life.

Kant applied autonomy strictly to the moral realm, though it had political implications. Since then Western Philosophy has come to term the capacity to determine one's own moral codes as 'moral autonomy.' To moral autonomy have been added the concepts of "personal autonomy" (the capacity to determine one's own actions) and "political autonomy" (the capacity to make one's own political decisions and have them heeded) (Dryden 2015; J. Anderson and Christman 2005). What these three definitions share is the concept of "self-governance." ICT's threaten human autonomy whenever they interfere with this self-governance. This occurs whenever ICT's make decisions for people, especially when an ICT imposes on the user a way of doing something which is antithetical to the manner in which the user would have chosen. Some threats to autonomy are universal under all definitions of autonomy, but most depend on the particular form of autonomy being used. Reducing autonomy may be justified on other grounds but such a justification, like the threat, will be based on the version of autonomy being used. It will often be the case that a justification works under one concept of autonomy, but not under others.

6.1.6. ETICA's technology groups and their threats to autonomy

As described above, the ETICA project identified the following eleven technology groups:

1. Affective computing
2. Ambient intelligence

3. Artificial intelligence
4. Bioelectronics
5. Cloud computing
6. Future internet
7. Human/machine symbiosis
8. Neuroelectronics
9. Quantum computing
10. Robotics
11. Virtual/augmented reality

Not all technology groups were seen by ETICA as generating ethical threats in and of themselves. ETICA assessed quantum computing as unlikely to have any impact on society in the next 10 – 20 years, while cloud computing and future internet were seen as enablers of other technologies and not as raising unique ethical issues themselves. We shall now briefly examine the remaining technology groups and discuss some of the ways in which each has the potential to reduce human autonomy. Our aim is not to explore the details of the individual threats, but simply to see how autonomy can be threatened by each technology. While a few technologies threaten autonomy no matter how it is defined, the central point of this survey is to demonstrate that most threats only appear under specific definitions of autonomy.

6.1.6.1. Affective Computing

Affective computing involves treating human emotion as input and generating output which either emulates human signals of emotion or which seeks to manipulate human emotions. Here ethical concerns derive from the importance of human emotion in communication and the risks of manipulation and misunderstanding.

How affective computing threatens autonomy can depend on whether affective computing works or not. Unlike some other technologies, there is no guarantee it will ever be possible to build systems which can accurately read human emotions (Barrett, Gendron, and Huang 2009; Picard 2003). The greatest danger with affective computing lies in the possibility that it people will think it is working when it is not. When most technologies operate incorrectly it is obvious. However, there is no simple way of checking to see if an affective computing system's assessment of a person's emotion is accurate. Incorrect assessment of someone's emotions threaten their autonomy when that data is used to make decisions which affect them.

If ICT's can successfully understand human emotions, affective computing threatens privacy through the ability to glean information about people which they do not wish revealed. Wherever such information is used to make decisions affecting the person, it may constitute a restriction of their autonomy. Here it depends on the form of autonomy one subscribes to. Some "second-order" procedural accounts of autonomy hold that autonomy is preserved provided one would have agreed with the decision if they had been given the chance (Dworkin 1988, 2015). However, accounts of autonomy which do not allow for such "second-order" assessments hold that any decision denied the person is a restriction on their autonomy. If those decisions were made in an effort to benefit the person, they *may* be justifiable on the grounds of form of paternalism. Otherwise they simply constitute a form of oppression.

Affective computing therefore constitutes a threat to autonomy whenever affective data is used to make decisions about someone, unless one is using a second-order procedural definition of autonomy *and* the person would have agreed with the decision if they had been given the chance.

Emulation of emotion makes possible threats to autonomy in that it offers the chance to overwhelm rationality with emotion. Any manipulation of a person's emotions which prompts them to make a decision they would not have made otherwise constitutes a restriction of autonomy, irrespective of how autonomy is conceived. The concern is therefore that introducing affective output changes the power balance between system and user. If the ability to accurately detect emotions is combined with the ability to emulate emotional expression, the power of such systems to manipulate and persuade becomes greater than any previous human invention. In this sense, affective systems become human manipulation technology.

6.1.6.2. Ambient Intelligence

Ambient intelligence refers to IC technology which is embedded into the environment. Its defining characteristic is the invisibility of the devices, often associated with automated input (for example, a change in room temperature) and less-than obvious responses by the ambient system (for example, changing the air conditioning settings).

Ambient intelligence offers the capability of personalising the environment to the individual user. Where personalisation is under control of the user, it represents an extension of their autonomy. Where personalisation is outside the user's control, it represents a reduction in autonomy under most, but not all, definitions. Second-order accounts allow for the preservation of autonomy if the user would have agreed to the personalisation had they been given the chance. Only under such accounts is autonomy preserved by automated environmental personalisation. If the personalisation is intended to improve the user's quality of life, reductions in autonomy may be justifiable on the grounds of paternalism. However, that is a problematic debate which cannot be resolved in terms of universal principles, but will ride on the specifics of each case; the type of personalisation, the individuals involved and their personal values.

Ambient intelligence shares several ethical issues with affective computing. Autonomy is threatened by poor quality personalization services, either as a result of an inadequate understanding of human nature, or by the skewing of personalisation protocols resulting from commercial interest or developer ignorance (or bias). Inference of user needs from their behaviour is particularly problematic, especially with regard to children. There is the danger that users will be forced to change their behaviour to get the best out of ambient systems, and so end up being "trained" by their environment's designers (Soraker and Brey 2007).

Ambient intelligence creates the risk of the personal environment acquiring power over people. Since ambient intelligence makes the personal environment controllable by third parties, the personal environment becomes a contestable zone; open to commercial and state interests, both in terms of gathering information and seeking control. Some contestation of the personal environment has always occurred. For example, disputes over noise can be traced back to the earliest cities (Goldsmith 2012). However, ambient intelligence represents such a significant increase in power that contestation for the personal environment becomes a new type of issue – the power to (potentially) totally control someone else's personal space.

6.1.6.3. Artificial Intelligence

ETICA examined both "hard" AI (emulation of human thought) and "soft AI" (e.g.: expert systems and software agents). Ethical concerns focused on the use of soft AI as an enabler for other emerging technologies, such as ambient intelligence and affective computing. While issues associated with hard AI were considered, ETICA's judged that hard AI was unlikely to do more than appear in earliest prototypes by 2030 and so was outside the scope of the project.

Our discussions of ambient intelligence and affective computing have both shown how human autonomy can be threatened by the power of these systems over people. AI significantly enhances this power, and so can be an enabler of threats to autonomy wherever it is deployed within ICT systems. Particularly problematic is the possibility that AI may enable systems to learn things about people not anticipated by designer or user. Implementation of informed consent becomes extremely difficult when one does not know what information may be acquired. It has been suggested the solution is to build ethical reasoning into AI systems (Dennis et al. 2013). However, ethical systems are abstract principles of value, not problem-solving algorithms (Kraut 2016; Graham 2004). Ethical dilemmas only exist where multiple ethical values conflict because it is these conflicting values themselves which create the dilemma. AI developers are not, therefore, able to reach for an existing corpus of ethical problem-solving algorithms as if the matter were a simple sort task. Any implementation of ethical processing will therefore impose someone's particular ethical values on the system. Should such an ethical value set not accord with the user's set of ethical values, the user's autonomy will be compromised by the actions of their AI-enabled system. The only way to counter such a threat would be to allow users to customise their AI-enabled systems with their own ethical values and reasoning.

6.1.6.4. Bioelectronics

ETICA defined bioelectronics as ICT systems which interact directly with the human body. The primary ethical considerations were focused on use in health care, but those who see bioelectronics as a path to artificial human bodies were also considered. Here bioelectronics represents the potential for technological modification of the human form. It therefore directly confronts the essence of what it is to be human and impacts core human values, such as autonomy, freedom and dignity.

Bioelectronics can be used to change people's internal states by delivering medicines and intervening in bodily processes and thus has the potential to reduce autonomy. However, this concern is dependent on the concept of autonomy used. If one accepts a second-order procedural concept of autonomy (Dworkin 1988), autonomy may be preserved *post-hoc* by agreeing with the earlier bioelectronic intervention. However, conceptions which situate autonomy at the moment of decision must see any bioelectronic intervention as a reduction of autonomy. Whether this loss of autonomy can be justified on paternalistic grounds depends on the intent behind the intervention.

However, it may be possible to reframe the issue such that autonomy is not threatened. Discussions of bioelectronics tend to treat intervention as morally equivalent to an intentional act. Under this perspective, each individual bioelectronic intervention is treated as a discrete act to which assent may be given or withheld. This raises problems for the preservation of autonomy because of the requirement for ongoing consent (Agich 2003). However, where this intervention is frequent and automated, such as is the case with drug delivery, it may be more productive to think of it as a form of reflex. Some definitions of autonomy incorporate reflex into their schema as non-cognitive expressions of the self (Meyers 2005, 1989b). Treating bioelectronic interventions as artificial reflexes removes any threat to autonomy under such definitions.

The nature of bioelectronic intervention means it is difficult to resist. This places a great deal of potential power in the hands of those controlling the devices; granting them the ability to control someone to a degree not possible by other means. The capability of bioelectronics to change internal bodily states and processes is the same as that of drugs. Accordingly, bioelectronics shares many of the same issues pertaining to autonomy as do drugs. There are today debates about the advisability of treating some mental states as illnesses, concerns about inappropriate use of drugs to control behaviour in elderly and children, and other issues relating to appropriate boundaries to medical intervention (B. Smith 2012). These same issues apply to interventions by bioelectronic devices. Threats to autonomy apply here in two ways; firstly, "authentic" conceptions of autonomy involve the concept of authenticity (that there is some given essential nature to each individual), such that autonomy requires being in conformance to that essential self (Kühler and Jelinek 2013). These accounts hold that there is something innate about humans which cannot be changed without harming their authenticity, such as particular forms of cognition (Frankfurt 1971). Theological accounts of autonomy may also depend upon a divinely-ordained authentic human nature (E. O. Wilson 1978; Niebuhr 2004). Meanwhile, "coherentist" accounts base autonomy on continuity with personal history, such that changes or decisions which radically depart from someone's previous patterns or tastes are inauthentic, and thus reductions in autonomy (Ekstrom 1993; B. L. Miller 1981). Bioelectronic interventions which take a person away from an essential self, however that self is conceived, are reductions to these forms of autonomy. Finally, where bioelectronic interventions are unwelcome and forced on the individual, they constitute a clear reduction in autonomy under all definitions of autonomy.

6.1.6.5. Human-Machine Symbiosis

ETICA characterised human-machine symbiosis as the pairing of innate human capabilities with ICT. This definition covers an extremely wide range of systems, including haptic interfaces, decision-support systems, computer-assisted surgery, augmented reality and direct neural interfaces. There is considerable overlap with other technologies, especially bio- and neuroelectronics and artificial intelligence.

Attempts to enhance human beings through the addition of ICT components may constitute reductions in autonomy under the authenticist views just outlined. Therapeutic devices, designed to replace lost human capabilities rather than enhance them, may be autonomy-preserving, but other accounts can hold that even these constitute a loss of autonomy (Bublitz and Merkel 2009). This is particularly the case where those devices require monitoring or control by others. A device may therefore reduce autonomy when it is operated by someone else, but enhance autonomy where it is controlled by the user (Sharon 2017). Such concerns are not limited to devices implanted into people's bodies. Use of external devices, for example, as aids to memory, may constitute reductions to autonomy, especially under accounts based on the concept of an "extended self" (Olson 2011; Rachlin and Jones 2010), in which parts of our personal identity are embedded in external objects, such as clothing or mobile phones (Ahuvia 2005; Belk 1988; Turkle 2011). For someone who embeds part of their identity in their phone, even something as simple as an enforced software update may constitute a reduction in their autonomy.

6.1.6.6. Neuroelectronics

ETICA defined neuroelectronics as technology interfacing between the human nervous system and electronic devices. While neuroelectronics clearly overlaps with bioelectronics, ETICA felt the intimate connection between the person and their brain meant neuroelectronics could not be adequately examined within a general treatment of bioelectronics.

As with bioelectronics, autonomy is not threatened when neuroelectronics is used to gather data, but only when used to induce changes. Since changes cannot be made without knowledge of internal states, gathering data is, however, a privacy concern as an enabler of reductions to autonomy. Any neuroelectronic system which produces changes within the person constitutes a threat under most accounts of autonomy. While this is clearly the case with conceptions of autonomy which involve an authentic nature or coherent life-history, it may also be the case under second-order procedural accounts (which allow for autonomy preservation if someone agrees to a change afterwards). This is because a person's agreement to a previous procedure may be merely the result of the changes they have undergone. Where these devices are installed to deal with cognitive impairment, such as dementia, informed consent is impossible and thus autonomy must be reduced (though this may be justifiable on paternalistic grounds). As with bioelectronics and human-machine symbiosis, any accounts of autonomy which involve a divinely-derived authentic human nature will view all neuroelectronic changes as reductions in autonomy.

6.1.6.7. Robotics

ETICA defined robots as "machines with motor function that are able to perceive their environment and operate autonomously" (Ikonen et al. 2010, 114). ETICA focused on new developments which increase the mobility and intelligence of robotic devices, permitting their deployment into wider areas of society, such as the home and healthcare, and focused on robots built for specific roles within these contexts. ETICA did not consider general-purpose, fully mobile devices controlled by strong artificial intelligence seeking to make war on humanity.

The degree to which robotics can threaten autonomy depends on the definition of autonomy used and the use made of the robot. As has been noted earlier, traditional procedural accounts of autonomy as self-determination have been criticised for unrealistically portraying people as isolated individuals and leaving no space in the account for communitarian elements such as the influence of family or culture (Stoljar and MacKenzie 2000; Stoljar 2015). On this basis some have gone so far as to argue autonomy is an unattainable ideal (Strawson 1994). A more common response has been to develop a treatment of autonomy which includes space for the influence of others, such as the concept of a *social self* (Meyers 2005, 44). The central premise in relational accounts of autonomy is the necessary role of other people in the development of one's values and the centrality of interpersonal relationships to human existence; that it is not possible to function autonomously without the influence of others (Stoljar and MacKenzie 2000; Stoljar 2015; Donchin 2000). Such conceptions of autonomy are threatened when robots replace humans in roles which have social externalities, such as health care and education. While robots are able to undertake the central tasks, the loss of the human contact constitutes a reduction in relational autonomy. The consequence of this is that there may be some tasks for which robots are not ethically suitable, for no other reason than that they are robots. If the social externalities of a work role are essential for the maintenance of the recipient's autonomy, then that role must be reserved for humans. Were we to replace nurses or teachers with robots, we could thus see people argue they have an inalienable human right to refuse to be served by a robot and to demand human service.

6.1.6.8. Virtual/Augmented Reality

ETICA's union of virtual reality with augmented reality into one technology group was based on the common feature they share - the imposition of digital output onto the human sensory field. ETICA defined virtual reality as occurring where digital output completely replaced sensory data. Where it did not completely replace external sensory data, ETICA used the term 'augmented reality' (Heersmink, van den Hoven, and Timmermans 2010, 114). ETICA's concerns were founded on the fact that such systems mediate or replace interaction with the physical environment.

Autonomy is threatened when virtual or augmented realities depict objects, people and places in the real world and thus become involved with the many ethical issues associated with depiction, cultural bias and its influence the development of attitudes and taste (Zimbardo and Leippe 1991). The imposition of foreign cultural models in virtual realities would constitute a threat to relational and social conceptions of autonomy (Tomlinson 1991), while lack of alternatives and lack of configuration options threatens all conceptions of autonomy.

6.1.7. Summary

Many of the technologies examined provide personalised services tailored to the individual. These technologies threaten human autonomy when they fail to personalise effectively. This may result from misreading the user, insufficient granularity within the personalisation, lack of user control, inadequate modelling of the user or by imposing on the user ways of living which are contrary to their values. A number of the technologies change power balances within society, granting for the first time, or greatly increasing, the actant power of the environment over the individual. This power threatens autonomy if used to impose on the user or to enable autonomy-limiting personalisation. Some conceptions of autonomy contain elements which can be disrupted by some ICT's. Conceptions of autonomy which include inter-personal relationships are challenged by robotics, while conceptions of autonomy which make reference to authenticity, coherent life-history, or some form of given human nature, are challenged by bio- and neuro-electronics and may also be challenged by human-machine symbiosis.

Irrespective of the nature of autonomy used, our survey of ETICA's research suggests there are certain characteristics any ICT is likely to have if it threatens autonomy:

1. **Surveillance:** Surveillance consists of the obtaining of information pertaining to an individual from their behaviour or communication, followed by the use of that data for purposes, either unknown to that individual or against their wishes (Lyon, Ball, and Haggerty 2012). While it can be used in a paternalistic fashion to enhance the individual's autonomy, our concern is when it is used in a manner which limits the individual's autonomy. The presence of a surveillance system is a necessary precondition for the use of personal information by others and for personalisation services because it constitutes the means by which that information supporting personalisation is gathered. Control of the outflowing data by the individual is therefore an effective way in which to retain autonomy. Thus control of privacy is an enabling right to autonomy.
2. **Disparity of Control:** Disparity of control occurs when a third party has a greater control over the use of one's personal data and the autonomy-limiting processes than oneself. This lack of control can be accomplished through lack of knowledge on the part of the individual as to what is occurring, through "take it or leave it" terms of use combined with a lack of alternatives, and through the lack (or concealment) of user configuration capabilities. In some cases, such as data profiling, not just functions, but entire industries may be concealed (Dainow 2015a; Federal Trade Commission 2014; Turow 2011).
3. **Insufficient Configurability:** Autonomy can be threatened by lack of variability in configuration options suitable to reflect the variations in the user's desires, style of operation or range of outcomes. This lack of variation can stem from:
    - A lack of recognition by the ICT developers of the need for, or even existence of, such variations.
    - Insufficient consideration of need for user variability during the design process.

- Insufficient granularity of configuration options.
- Systemic biases within the delivery mechanisms, including business models, market competition, regulatory framework or any other factors regarding the operational delivery of services. For example, a system might be highly configurable, but delivered to users with a standardised configuration which cannot be changed in an effort to reduce support costs.

4. **Insufficient variation in operational models:** Many new ICT's are owned and operated by a very limited number of providers (Noam et al. 2003; Hillis, Jarrett, and Petit 2013). Often a single provider dominates the market to near-monopoly levels. This limits choice of service provision and consequently the model under which it is provided. In many cases the business model is identical across all providers within an ICT sector, such that change of provider does not change the circumstances under which the user accesses the service. For example, social networking is only available under a capitalist for-profit model in which service provision is exchanged for personal data which is then commoditised (or monetised). Non-profit and privacy-preserving social network systems do not exist to the degree that a real choice is available. In addition, IP protection promotes walled gardens and suppresses interoperability, further limiting options for choice by users because it is not possible to interoperate between social network providers.

6.1.8. Conclusions

The common feature to all conceptions of autonomy is self-determination. What one self-decides, how, on what basis, according to what procedure, and subject to which influences, varies from account to account. Yet efforts to preserve autonomy are central to ICT ethics (Spiekermann 2015). Efforts to reduce the negative ethical impact of ICT's on people through processes such as value-sensitive design need to be cognisant of the various conceptions of autonomy relevant to the intended functionality of the system. This is especially the case with technologies which enter into spheres of life which have previously been purely human or which reduce the power people have over their personal lifeworld. Given the wide range of definitions of autonomy, no single definition can possibly cover all cases.

Because autonomy is frequently central to the ethical status of any technology, it is rarely possible to categorise any ICT as threatening or reducing autonomy without reference to the form of autonomy used. Methodologies to improve the ethical sensitivity of ICT technology, such as value-sensitive design, make frequent reference to the need to specify a particular brand of ethics, such as deontology or utilitarianism, from which to draw values (Spiekermann 2015; Nissenbaum 2001; Winfield, Blum, and Liu 2014), but the application of such values still depends on the concept of autonomy being used. It is generally assumed that there is a single, fixed meaning to autonomy, at least in any given context, and that preservation of autonomy simply requires, at most, identification of the correct version (Manders-Huits 2011). Some have noted the meaning of autonomy can vary with different contexts (Burmeister 2016), but it is possible that it only appears the meaning has changed because the difficulties perceived apply to one version of autonomy and not others, thus making the problematic version of autonomy obvious. Our review of ETICA's survey of threats to autonomy has revealed that many versions of autonomy may be applicable within the one context, some of which may have problems while others do not. Thus, attempts to find the "correct" version of autonomy for any context may be moot. On this basis, it seems more appropriate to develop ICT functionality which does not depend on a single definition of autonomy, but adapts to the user's own definition. This implies a loose pairing of system behaviour to coding, such that output can be fine-tuned after deployment, either by a user, their technical support staff or through the use of soft AI components focused on adapting the system to user feedback.

All significant emerging ICTs can threaten autonomy, but there is no simple way of avoiding this. Efforts such as value-sensitive design stand more chance of success if they focus less on incorporating a single set of values and focus instead on ways by which users may adapt ICT's to reflect their own individual values.

-end of published paper -

6.2. Next Steps

The published paper answered the primary research question of how emerging ICT's could restrict autonomy. In and of itself, it is useful for evaluating specific technologies against the different versions of autonomy. However, while it goes some way towards developing general characteristics from which rules and principles can be drawn, it does not consider the implications of the fact people do not use only one ICT technology, but many, and that the impact of any one technology may be influenced by others. For example, the impact of robotics cannot be understood without reference to artificial intelligence. This tendency to focus on a single technology, without reference to the technological context in which it occurs, is a common characteristic of ICT ethical analysis. We therefore continued our research by looking for existing analysis which *did* consider the impact of combined technologies within the context of emergent ICT's. Our research indicated the best framework by which the future state of these interacting ICT's can be analysed was smart city research. While there are many conceptions of the smart city, it is, by definition, a descriptive conception of a state of affairs produced by the interaction of many ICT's and people and forming a complete environment.

In order to evaluate the combined effects of all emerging ICT's on individuals we therefore turned to smart city research. This first focused on the concept itself, which is typically considered within Geography, Architecture and related disciplines (such as Urban Planning) and then on more detailed discussions within computer science regarding specific technologies (such as city dashboards, expert systems in urban management, ad hoc networking, autonomous vehicles and IoT devices). It became clear there was a lack of models in computer science which accommodated both the complex technical infrastructure and human practice. In general, computer science accounts which did try to create comprehensive models focused on the technology, especially its formal and material aspects, while accounts from Geography, Architecture and related disciplines tended to focus on the social aspects, primarily in terms of teleology. However, an ethical account within computer science requires both in equal measure. It was therefore determined that a descriptive framework which could accommodate both the engineering and the human as equally important elements was required which could then serve as the platform for analysis of restrictions to autonomy within a generic digital environment (the smart city). This analysis stretches over the final two publications. The first publication elucidates the descriptive framework we developed, while the second explores the manner in which ICT's may restrict human autonomy within this framework. The descriptive framework was published in 2017 in *Orbit (1)* as "Smart City Transcendent". The analysis is found in the book chapter, "Binding the Smart City Human-Digital System with Communicative Processes", to be published in *Technology and the City* (in press).

Our research at this point focused on the concept of a digital environment. A digital environment is one in which ICT's and other digital technologies are the dominating defining characteristics, such that significant elements of one's life are unavoidably mediated through digital technologies. Smart cities are visions of such a digital environment, though there may be other ways of envisioning it. Our interest in the smart city is not motivated by a belief that smart cities, as envisioned today, will come into being. Whether today's predictions come into being is irrelevant. The reason for considering smart cities is that there a rich source of literature available for research purposes, covering concepts, scope, disciplines, methodologies, elucidating issues and contextualising technology innovation. Furthermore, this research material is enhanced by empirical data regarding current smart city initiatives. Most importantly, these smart city initiatives mean smart cities meet the same criteria for socially significant and emergent technologies used by ETICA. By evaluating digital environments through the lens of the smart city, we thus maintain methodological consistency. The smart city thus constitutes a legitimate expression of a situation in which ETICA's ICT's operate simultaneously, while also allowing for (unspecified) new social practices in their use.

7. Understanding the Smart City as a System

The paper presented in this chapter provides a conception of the digital environment by considering what the smart city brings to the world which is new and original. This approach provides a means of dealing with the complex influences humans and digital systems will have on each other in the mature smart cities of the future. The paper first reviews traditional accounts of the smart city and derives from them the essential characteristics common to these visions. We then show how these characteristics can be best understood through Actor-Network Theory (Latour 2011) and construct an account of the smart city as an autopoietic system in which humans and devices are co-constituting actants. Finally the paper develops this into an original conception of the smart city as a new type of thing - an *Integrated Domain.*

The following section contains the paper as published. Minor modifications have been made to ensure the writing style matches that of this thesis (such as changing "I conclude" to "we conclude"). The original published paper used the term 'communicative triad.' However, this term was changed to 'integrated node' in later accounts of the theory, and so has been changed here from the original 'communicative triad' to 'integrated node.' Headings have been numbered and the citation style has been changed to match the rest of this thesis.

7.1. Smart City Transcendent - Understanding the Smart City by Transcending Ontology

Orbit 1 (2017). https://www.orbit-rri.org/volume-one/smart-city-transcendent/

7.1.1. Accounts of Smart Cities

There is no single definition of what constitutes a smart city. Accounts divide into three schools regarding definitions of the smart city. One school defines the smart city in terms of what we can do with it. Here we find topics such as innovation policy (Komninos, Schaffers, and Pallot 2011), smart governance (Vinod Kumar 2015), urban planning (Zygiaris 2013), improved sustainability (Bowerman et al. 2000) and similar large-scale management tasks. Others define the smart city in terms of its material construction. Here the focus is mainly on devices comprising the Internet of Things and their interactions. Some focus on issues of designing smart sensors and other components, such as weight detectors in roads (Hancke, Silva, and Hancke 2013). Others focus on the interactions between components, such as sensor networks

(Filipponi et al. 2010) and human-sensor interactions (Pettersson et al. 2011). Others attempt to make sense of both usage and components through the development of organisational models. For example, Jin proposes a four-layer model of sensors, networks, cloud-based data management and service delivery (Jin et al. 2014). Organising smart city components and operation into layers is fairly common. Balakrishna's four-layer model (Balakrishna 2012) is almost identical to Jin's, while others adopt a three-layer (Atzori et al. 2012) or five-layer (Pettersson et al. 2011) model. In contrast, some have attempted more conceptual organisational models. For example, Filipponi (Filipponi et al. 2010) suggests a vertical model based on our understanding of the space in which activity occurs, such as car, home, office and civic space.

Attempting to make sense of this range, *Smart Cities: Definitions, Dimensions, Performance, and Initiatives* (Albino, Berardi, and Dangelico 2015) examined twenty-two accounts and cross-referenced these with urban developments which labelled themselves as smart cities. Their conclusion was that a single definition was impossible because of wide variations in the meaning of common terms. Attempts to include everything inevitably cover too much or too little. At one extreme we have very broad statements, such as a smart city is that which uses "social, mobile and sensor-based technologies … to create more productive alignments between (growing) demand and (constrained) resources." (Hartswood et al. 2014, 3). At the other extreme, *Getting Smarter About Smart Cities* defines a smart city as:

> "using networked, digital technologies and urban big data to tackle a range of issues, such as improving governance and service delivery, creating more resilient critical infrastructure, growing the local economy, becoming more sustainable, producing better mobility, gaining transparency and accountability, enhancing quality of life, and increasing safety and security." (Kitchin 2016, 9).

Deployment of smart city technology is more likely to be accomplished by industry than by academic researchers. It is therefore worth considering the vision of the smart city which industry offers. Cobham Plc is a major player in the smart city sector, with annual turnover of $US3 billion (Cobham Plc 2011). Their Tactical Communications and Surveillance division provides a range of smart city technologies. Their smart city sales brochure ('Safe Cities' 2014) offers a 5-layer model, ranging from an "IP mesh" which allows any type of sensor to communicate with any other, through several layers of device type, including integration of personal devices as sensors (with and without user knowledge), through to management systems.

Certain characteristics are held in common throughout all these accounts. They agree that smart cities require a ubiquitous heterogeneous sensor network which provides information about the inhabitants, their environment and service delivery. This has been referred to as an “IP mesh” (Cobham Plc 2011, 3) and as an “underlying sensor fabric” (Balakrishna 2012, 224). These sensors must report their data to other devices. Some of these communication patterns can be predesigned, while others must be created on an ad hoc basis in response to the movement of the inhabitants and machines such as drones, robots and cars (Guo et al. 2013; Pettersson et al. 2011). There is general agreement these devices will be embedded in the civic environment, the home and other personal spaces (such as the car), worn on the person and implanted within the body. While many accounts assume that all data processing will occur in the cloud, this is not inevitable. A contrasting view can be seen in the concepts of fog computing (Bonomi et al. 2012; Petrolo, Loscrì, and Mitton 2015) and user-controlled personal data stores (Service Systems Group 2015b). Under these views significant data processing can be done locally, either within sensor devices themselves or through locally situated data processing units (Service Systems Group 2015a). Indeed, such local processing is likely to be more secure and efficient (Dainow 2015b; Langheinrich 2001).

The aim of this paper to provide a comprehensive definition of the smart city which can serve as the foundation for ethical analysis by reconceptualising the smart city. Current ethical concerns for the smart city tend to be limited to specific issues associated with specific technologies within the smart city, such as cars (Jaisingh, El-Khatib, and Akalu 2016) or location detection (Martínez-Ballesté, Pérez-Martínez, and Solanas 2013). Our aim is to provide a comprehensive framework for analysis of any ethical issue, especially issues which are emergent from the interaction of multiple systems and which therefore cannot be predicted from any one sub-system. This requires developing a new vision of the smart city which is able to account for the complex nature of interactions within it.

7.1.2. The City of the Future

We must first place the smart city within a foresight studies framework. This is because the smart city does not yet exist, but is rather at an early stage of development. Given that the smart city does not yet exist, current technologies related to smart cities must be regarded as interim steps towards what will be the ultimate deployed solutions in mature smart cities of the future. We must therefore regard current smart city technologies as in flux, as unfinished, and their features as mutable and almost certainly subject to change. Furthermore, our current understanding of the path to the mature smart city is highly uncertain. The issue of ad hoc networking illustrates that even the direction of innovation is in dispute. Much concern in smart city development focuses on issues of communication between devices. It is widely recognised that mobile devices, autonomous systems and moving people will all necessitate the dynamic creation and uncreation of unpredictable network patterns. Known as "ad hoc networking", some see great advantages in this (Boldrini, Conti, Delmastro, & Passarella, 2010; Guo, Wang, Zhang, Yu, & Zhou, 2013) seeking only to make it reliable and secure, while others (Pettersson et al. 2011) see its chaotic and unpredictable nature as something which needs to be controlled. Clearly, it is difficult to anticipate the ultimate role of ad hoc networking in the smart city at a stage when we cannot even agree on whether it should be promoted or inhibited.

Teleological accounts, concerned with urban governance, typically focus on the ways in which smart cities can solve today's problems. However, it is unlikely that the contentious issues of smart city governance are predictable in any detail. Prior to the invention of the world wide web, no one could have anticipated the need for domain name registries, nor the political fighting which would emerge around them. We can therefore anticipate that at least some important forms of governance of mature smart cities are yet to be imagined. Similarly the most pressing issues confronting mature smart cities will include those we cannot yet anticipate because they will derive from the unknown nature of currently unpredicted technologies and the currently unpredictable patterns of usage which will evolve around them. Considering that new technologies always create negative side-effects (Derry and Williams 1993; Cardwell 1994), it is inevitable that the solutions we deploy in creating the smart city will generate problems of their own – which we also cannot predict.

Placing the smart city within a futures perspective therefore means understanding that we do not know how it will be made, how it will operate, or how it will be managed. Conceptions of this future unknown smart city must therefore frame themselves in terms which are not dependent upon knowledge of the specific details of future smart technologies. Fundamental to the futures approach is an understanding that the smart city is not today's city with some digital tech laid over the top of it, any more than the modern city is just a Victorian one with cars instead of horses. The mature smart city will be a fundamentally different type of environment from any we have previously seen. The key change will be the presence of ubiquitous ICT technologies. Through all of history the human built environment has been a largely dumb. It has not had the ability to do things to us except in the most gross fashion and it has not had the capability to know us. The capacity of the environment to obtain data and to respond to human action represents a fundamentally new type of built environment for human living.

We can therefore understand the characteristics of the smart city best if we consider it as a form of built environment. Accounts of smart cities focused on the technical infrastructure typically treat it as consisting of intelligent devices embedded within dumb materials. However, the number, ubiquity, heterogeneity and invisibility of most devices will cause humans to understand and act towards the object in which the devices are embedded rather than the devices themselves, and in many cases deal with aggregations of devices rather than individual ones. Attempting to account for this interaction between human and device in terms of component types and processes is impossible on two grounds. Firstly, we cannot know the nature of these future devices or what patterns of usage humans will develop towards them. Secondly, the number and heterogeneity of these devices and variability of human relations with them across differing contexts renders attempts to understand on the basis of technical type impossibly complex. We must therefore cease to talk in terms of technical components and instead identify the functional characteristics which can unite the various heterogeneous technical forms.

The following will identify the essential functional characteristics of the smart city. Thereafter we will use these to synthesise a conception of the smart city as a new type of being.

7.1.3. Ambient Intelligence

Ambient intelligence is defined as technology which embeds input, processing or response ubiquitously through the environment (Ikonen et al. 2010). With its ubiquitous sensor mesh the smart city is the ideal type for ambient intelligence. As a lived experience, people will know the smart city not as individual components and discrete processes isolated within atomised sections of their lives. Instead the smart city will be experienced as a seamless experience as one moves from house to car to work, from one context to another. People will not think of themselves as moving from one discrete situation to another, they will simply experience themselves as going about their daily life, an unbroken stream of changing contexts seamlessly merging from one to the other. The following quotes from the auto industry illustrate current steps in this direction:

Matt Jones, Director of Future Technology at Jaguar Land Rover:

> "We very quickly discovered that customer expectations really aren't based on our Jaguar Land Rover competition - they're based on the smartphone. They're based on the tablet. They're based on the home entertainment experience…. Consumers expect to be able to download an app, get into the car and discover a seamless integration between the device and the automobile." (Bedigian 2016, 7)

John Schnoes, Programme Director of Vehicle Information Technology at Nissan:

> "It all comes down to smart device connectivity. People are really looking to have that sort of seamless experience when they've been working on their phone and then they walk into their car and they expect the platform to know what they've been doing." (Bedigian 2016, 8)

Emergent in-car systems reveal that cars will communicate with their environment in a deep fashion. Usage-Based Insurance systems analyse driver behaviour, including time of day, acceleration rates, cornering and locations travelled in order to dynamically set insurance premiums. Emerging collision-detection systems can automatically contact emergency services with essential details, including the identities of the occupants and which seatbelts were fastened. Car maintenance systems will access user calendars to determine the best time for maintenance appointments. Already today, the Ford SyNC systems can accesses the driver's medical data to remind them when to take their medicine (Jaisingh, El-Khatib, and Akalu 2016).

We can see similar cross-context connectivity trends at the urban management level. For example, Restore NV provides demand management systems, such as those controlling electrical grids, including five of Europe's largest grid operators (Restore NV 2016). Their newly announced smart city electricity management system, FlexPond, monitors individual home appliances to predict electrical load within individual homes (Restore NV 2017). This information can be cross-referenced with local weather patterns and the number and type of household occupants to greatly improve demand prediction (Hancke, Silva, and Hancke 2013).

Other smart city systems treat personal digital devices as mobile components of the smart city infrastructure. This is not limited to using such devices merely as sensors or as personal identifiers. Patterns and content of social interaction between people can be analysed to anticipate movement patterns, shared use of resources, and even as a way of determining appropriate security levels to create between devices (Atzori et al. 2012). Others have tested analysis of social networking to guide structuring of ad hoc networks (Boldrini et al. 2010). China plans to take this further, creating a "social credit system" by monitoring many aspects of personal conduct, including honesty and conformance to socialist behaviour. A key aim is to reduce social diversity. This system will offer rewards and punishments through differential access to smart city services (State Council of People's Republic of China 2014). Some smart city services will even penetrate the physical body. Already being tested, body sensor networks embedded into the environment communicate with implanted medical devices, such as heart monitors, to anticipate and react to medical emergencies (Lo et al. 2005). Such health sensor networks can also monitor some vital signs externally, such as skin temperature, respiration, sleep patterns and diet (Hancke, Silva, and Hancke 2013; Y. Kim, Kim, and Lee 2008).

The picture which emerges is one in which smart city services must take data from a ubiquitous digital ecosystem, in which digital devices are embedded throughout the fabric of the built environment - "transforming everyday objects into information appliances," (Botta et al. 2016, 691) but also including devices carried by people and embedded within their bodies (Filipponi et al. 2010; Balakrishna 2012; Botta et al. 2016; Jin et al. 2014; Pettersson et al. 2011). This data will be integrated across contexts, technology forms and purposes so as to create an integrated sensing and response environment (Psyllidis 2015). Such an "intelligent information infrastructure" (ITU-T 2008, 2) is clearly an ideal type for ambient intelligence.

7.1.4. Artificial Intelligence

Central to the vision of the smart city is algorithmic intelligence, or "soft AI" (e.g.: expert systems and software agents) (Komninos 2006; Ikonen et al. 2010). It has been estimated that a typical smart city will contain around 1 trillion nanoscale devices (Balakrishna 2012). A central paradigm implied within the concept of the smart city is that the smart city will generate new data and novel possibilities for action as a result of the integration of data from disparate domains (Picon 2015). There is a general acceptance that this will require machine learning and other forms of soft AI (Balakrishna 2012; Nam and Pardo 2011). In addition to common expectations that soft AI will provide core functionality through cloud computing and big data, it has been suggested soft AI will need to be embedded at local level to support ad hoc networking and to facilitate more rapid system responses due to the sheer scale of data (Komninos, Schaffers, and Pallot 2011).

7.1.5. Robotics

Robotics refers to ICT devices possessing the ability to move autonomously (Ikonen et al. 2010). Self-driving cars (and possibly drones) are expected to exist within smart cities (Jaisingh, El-Khatib, and Akalu 2016; Balakrishna 2012; Petrolo, Loscrì, and Mitton 2015). The home will see robotic devices such as vacuum cleaners and similar devices, sometimes known as mobile IoT (Gubbi et al. 2013). The specific details of which robotic devices will exist within the smart city need not concern us for the purposes of this analysis. The important point is that the smart city environment will involve both humans and machines moving within the same spaces. Just as the introduction of the car eventually led to the rise of pedestrian crossings, so an environment of moving machines will require adjustments in human behaviour to take them into account.

7.1.6. The Smart City as a Socio-Technical System

The vision which has emerged of the smart city places technical artefacts and humans as equally powerful actants. Digital devices will communicate amongst themselves and engage in negotiations (Atzori et al. 2012; State Council of People's Republic of China 2014) as they create ad hoc networks, including event-driven networks (Filipponi et al. 2010), contextual networks (Boldrini et al. 2010) and social networks (Atzori et al. 2012; Boldrini et al. 2010). Thus the network structure of the smart city will be created by the devices as well as by the humans. Device activity will respond to human activity. Humans will respond to these responses. Furthermore, digital devices will have their own needs, which will require consideration by humans. Accordingly we can view digital devices as causal agents in a smart city on the same causal level as humans.

Systems theory has dominated urban planning since the 1960's (Taylor 2005) and permeates approaches to the smart city (Söderström, Paasche, and Klauser 2014), which is seen as a complex system, or system of systems (Albino, Berardi, and Dangelico 2015; Chourabi et al. 2012). Actor-Network Theory (ANT) (Latour 2005) provides a theoretical framework for incorporating the smart city's digital agents and their needs into an interactive relationship with humans. By treating both devices and humans as co-existing within the same structure, ANT provides a framework by which we can attribute to artefacts causative properties within the human dimension and vice versa (Tabak 2015). A further advantage of ANT is the lack of need to specify the details of each node within the network (Law 1992). This suits a both a heterogeneous device mix and a futures approach which accepts that we cannot know what final forms smart city technology will take. In addition, ANT has been used to account for pathways in ICT innovation (R. Kim and Kaplan 2005; Lamb and Kling 2003) and so applying it to smart cities can be linked to that research. This offers the potential for an explanatory framework which can account for the dynamics of the innovation which will lead to the mature smart city.

The arrival of an intelligent responsive environment, especially one containing autonomously mobile agents, requires changes which must affect people. This will inevitably bring the needs of the digital agent into conflict with the needs of the individual or society, just as cars require roads and regulations which gives them effective rights over people in some circumstances. Ambient intelligence systems may also require changes in human behaviour, raising the possibility that our environment will train us to suit its needs (Soraker and Brey 2007). In other cases, we can anticipate that personalisation of shared spaces will result in differences between people regarding how much any particular personalisation suits them. This may range from the trivial, such as ambient temperature, to the existential, such as access to sensor networks required for medical implants. As the Chinese social credit system shows, some people will be forced to make compromises in order to accommodate aspects of personalisation preferred by others within the shared space. It is not inevitable that spaces will be personalised to suit the majority, or that minority needs will be accommodated. It is possible that environmental personalisation will become a zone of contention and power dynamics. Given the potential impact alteration of spaces can have on the individual, control of personalisation of shared spaces could easily become a path to domination of others. Hence control of the digital actants within the smart city's actor-network can be expected to grant influence and power over the human actants.

The fact digital actants have their own needs which require changes by humans raises the issue of secondary rights. Digital objects, while requiring humans change, do not, in and of themselves, originate that need. They are operational units whose existence is caused by humans and to whom the benefit of their activity is given to humans. They are not self-originating and they are not in receipt of the benefit of their activity. As a result, while it appears that the object is competing with the person, in actual fact the object stands in proxy to another individual - the beneficial owner. While the situation may look like competition of rights between the individual and the object, it is in fact competition between one individual and another in which the device stands as proxy for the second individual. However, the rise of AI systems may mean that, while other humans benefit from the satisfaction of digital needs, they did not originate those needs. Instead we may encounter needs that were generated by the AI system as a result of its own development and self-learning. Given the expected complexity and scale of a smart city's ambient intelligence, there are likely to be many contexts in which we cannot distinguish between human-originated and self-determined needs of digital actants. It is likely that some needs will arise via a combination of human-originated and self-determined needs. It is likely that many human goals will be accomplished via methodologies which were self-determined by autonomous systems. Where a self-determined methodology has negative effects on other people, we can expect some form of resistance. Such a situation represents an interactive process between human and digital system, in which human decision-making contends with the autonomous decisions of a digital actant. In that ANT considers nodes as "black boxes" (Tabak 2015, 37) and does not need to specify their internal details, it is not confounded by the issue of whether an activity derives from the device or some beneficial owner.

By treating the human and the digital components of the smart city as equal actants within the same environment, ANT also offers a dimension of analysis which is essential for a full understanding of the interactions of the human and the digital – the perspective of the digital device. By its very nature, a digital device cannot know the physical world. What a device knows is what its sensors generated as input. For example, a thermostat does not respond to the temperature, but to what the sensors tell it the temperature is. If the sensor malfunctions, or we game it, the thermostat control will still respond to what the sensor tells it, not what the material reality is. Hence digital devices do not know the physical world, but inhabit a totally digital environment. What they know of us and how much they know is determined by our representation in digital terms. Smart city systems will not respond to us; they will respond to our digital representations. If those digital representations are incorrect or incomplete, then the analysis of us will be compromised and the delivery of services will suffer.

Accordingly, ANT offers us a conception of the smart city as a complex network of heterogeneous actants. Under this view we recognise that these actants may be a single digital device, an individual person, a group of people, a group of devices, or a mixture of both. We can build on this conception by considering the logic which must be inherent within operational smart cities in order for them to exist. Doing so will provide a more detailed understanding of the nature of the smart city system, revealing its essential quality of autopoiesis.

7.1.7. Autopoiesis in the Smart City

The concept of autopoiesis originates in theoretical biology in the 1970's (Varela, Maturana, and Uribe 1981) and accounts for living things, such as cells and bodies, as self-sustaining systems. The concept was made accessible to the social sciences by Luhmann (Luhmann 1986a) in the 1980's. Systems theory holds that the defining characteristics of a system are not determined by the properties of the components but by the patterns of their relationships (Varela, Maturana, and Uribe 1981). Autopoietic systems are those which are self-maintaining in the face of changing stimuli, either changes inside the system or within the system's environment. To be autopoietic, systems must also regenerate the essential patterns which characterise that system and internally generate the components and processes required to maintain it (Maturana 1981).

The original theories of autopoiesis were designed for biology and seemed limited to that field by the necessary characteristic that an autopoietic system generate its own components (Maturana 1981; Varela, Maturana, and Uribe 1981), a process known as "material self-production" (Di Paolo 2005, 433). However, this issue is really one of the minimum granularity in one's ontological schema. Living systems are the ideal type for theories of autopoiesis, but they do not create atoms or many of the molecules they depend on. No autopoietic system generates the materials from which the components considered necessary for characterisation of that system as autopoietic are created. All autopoietic systems exchange input and output with their environment (Luhmann 1986a). Hence autopoietic systems need only self-generate the components which are the most proximate cause of their characteristics and self-maintenance.

Building on this approach, Luhmann reworked the concept of autopoiesis to bring it into the social sciences (Luhmann 1986a). Under his account social systems are autopoietic systems which use communication as their characteristic form of autopoietic generation. The atoms from which social systems are constituted are communications, which are recursively produced and reproduced. Luhmann defines communications as being composed of three elements; information, utterance and understanding. As such, a communication is an event. A social system as autopoietic is a "network of events which produces itself … the reproduction of events by events." (Luhmann 1986a, 175). The critical point is that the elements of an autopoietic system which make it autopoietic are those which maintain autopoiesis in the face of destabilising forces and which give rise to characteristics of autopoiesis which distinguish it from other autopoietic systems. The basic unit of social autopoiesis is this integrated node of information, understanding, and utterance (Luhmann 1986a).

This pattern is easily visible within the smart city. It is the minimum necessary capability of digital device interactions which exhibit any of the defining attributes of an ambient intelligence. Within ambient intelligence this communicative pattern is visible as the ability to accept input ("information"), process it in some way ("understanding") and consequently produce output ("utterance"). In order for any device to participate in an ambient intelligence it must be able to perform these three processes. The actant within the autopoietic system is thus defined by its ability to accept input, process it and respond. The response becomes input for a connected node and thus the process perpetuates.

However, smart cities will not be constituted by closed communicative systems in which digital devices communicate only with each other. As we have seen, the smart city will be characterized by the close integration of digital devices and humans (Bicocchi et al. 2013). Much of the digital device's input will derive from humans, both as intentional commands and as unconscious input (such as movement and reactions to digital service delivery). Hence both devices and humans will react to each other and stimulate new responses in turn. In this way both humans and digital devices will be seen to engage in communicative patterns amongst their own kind and across the human-digital divide. Furthermore, once begun, this system becomes self-referential and recursive, as each responds to the response of the other, and so the system becomes self-sustaining and autopoietic.

If we accept this *integrated node* as the atomic unit of a living smart city, there exists a temptation to see this as two societies, one human and one digital, communicating within themselves and only interacting with the other by means of constituting its environment – humans embedded in a digital environment and digital devices surrounded by humans. However, significant portions of human communicative action will be mediated by the digital environment, while that digital environment's communicative patterns will respond in myriad ways to every human action. Under such circumstance, few integrated nodes can be said to be purely human or digital; the overall patterns of communicative flow will be the result of both human and digital components. Some have highlighted the tight interaction of the ICT and human communities in smart cities under the terms 'collective sensing', 'collective awareness' and 'collective action' (Bicocchi et al. 2013). However, such analysis does not incorporate the deep fusing of both collectives which must occur. We are not suggesting this fusion constitutes some form of hybrid human-machine society, or that we should redefine the term 'society' to include digital systems. Some have used the term 'smart society' to characterise the society within a smart city (Hartswood et al. 2014), but this refers to a human society using smart city services. This is a useful definition to maintain because it refers to real issues and to a clearly definable zone of concern. We therefore believe we need a new term for this new phenomena, one which refers to an autopoietic system constituted of integrated nodes which may be digital, human or both. We propose to call this an *integrated domain*.

7.1.8. Defining Integrated Domains

An integrated domain is a socio-technical system in which humans and digital devices co-mingle in a manner such that it becomes impossible (or even meaningless) to identify the origins of patterns within the system as being either human or digital. An integrated domain is autopoietic. The autopoiesis of an integrated domain derives from a close coupling of nodes such that output from one node cannot avoid generating a reaction in another node, combined with the unavoidable structure of the atomic node, which is the integrated node event. An integrated domain consists of two collectives integrated into a mutually dependant and co-creative partnership. One collective consists of a human smart society. This is a form of human society existing within an ambient digital environment, such that human perceptions, actions and intersubjectivity are unavoidably mediated and influenced by this ambient digital environment. The other (non-human) collective consists of an autopoietic system of digital devices and networks based on the integrated node. However, neither collective possesses strict boundaries against the other, but rather the two intermingle. We can thus describe each as separate for theoretical purposes, but it is not possible to account for characteristics of, or phenomena within, either without reference to nodes within the other. Any discussion of the nature and processes within smart society on the human side or the smart city digital components must draw on aspects of the other, and so any explanatory discussion of either must occur at the higher ontological level of the integrated domain.

The concept of autopoiesis includes a persistent demand for regeneration (Varela, Maturana, and Uribe 1981). This holds true for integrated domains as well. However, whereas in a living organism it is the material constituents which need regeneration, in an integrated domain it is the integrated nodes. Integrated nodes are events, not states (Luhmann 1986a). Patterns of communication thus form an actor-network in which the nodes are events constituted of integrated nodes, whose material base may be a human, a digital device, or a group of either or both. It is essential that groups of people and/or devices can be treated as equal actants within this model because groups will interact and communicate with individual devices or people in the smart city, and vice versa. An example of a combined group which would yet constitute a single node is a car in motion. The car is composed of a wide range of interacting devices. Much of the operation of these devices will be influenced by the actions of the driver, and vice versa. However, this complex system may communicate data aggregated from many of these systems in combination with the driver's actions to a single device, such as a traffic light. Similarly, a single device, such as a traffic light, can be expected to communicate with dozens, if not hundreds, of similarly complex entities. We thus see that the communicative system is integrated not just across differences of materiality, but also spans (or ignores) multiple ontological levels.

7.1.9. Social Machines

The primary unit of the integrated domain is the integrated node. This is a three-stage process consisting of input, processing and output. It constitutes an individual node within the network structure of the integrated domain. This network structure is dynamic, variable and self-sustaining. Under this view the smart city is seen as a system possessing autopoiesis. Each node is not a state, but a process or an event and their influence on the system's patterns is not determined by their material base. The material base of each node may be an individual processing unit, such as an single digital device or an individual person, or it may be constituted by a grouping of devices, people or even a combination of both. Each node may also be constituted by the aggregated activity of a group of sub-nodes. These sub-nodes may themselves may be constituted by another group, an overlapping set of groups or an individual person or device. It is likely that the cascade effect of events and responses, combined with the volume and complexity of communication will make it impossible, or even meaningless, to ask whether a feature of note is the product of one node or many, human or digital. The integrated node therefore constitutes both the atomic unit for, and the essential pattern of, interactions within the smart city. It provides us with a framework for handling the way in which smart city processes can transcend normally distinct ontological boundaries and frequently renders them meaningless. This deep interconnectedness of materially indeterminate nodes combines with ad hoc networking and human response variability to create a highly complex network structure.

As an autopoietic system, many of the integrated nodes will be concerned with internal states of operation within both digital assets and humans. In the case of digital assets, much of the communication can be expected to be concerned with issues such as maintenance and management of digital technology. This monitoring represents self-awareness. This is not awareness in the human sense, implying sentience. However, as a system, the integrated domain is aware of its internal processes. Through the component of the human, the integrated domain possesses the human level of understanding and meaning-giving. An integrated domain is therefore a self-aware, autopoietic system composed of actants who may be digital or human, individual or group.

We are already witnessing early examples of hybrid systems which combine human and digital devices under the label of "social machines". Social machines are defined as "socio-technical systems which involve the participation of human individuals and technological components…able to extend the reach of both human and machine intelligence [by] supporting capabilities that less integrated systems might find difficult to accomplish" (Smart, Simperl, and Shadbolt 2014, 55–56). Social machines are the early fore-runners of the integrated domain which will form the fabric of future smart cities.

7.1.10. Conclusions

The close integration of a huge range of devices and systems makes any model of the smart city which depends upon material differences impossibly complex and inevitably incomplete. The deep integration of various ontological levels, from the individual nano-sensor to the global cloud, makes it impossible to comprehend causes of individual actions within the smart city. Both of these issues are further magnified by the expectation and need of the human lived experience to transcend these technical distinctions. Discussion of smart cities in terms of enablers of human activity, such as urban management, do not adequately incorporate into their models the deep fusion of digital cognitive processes with human deliberations. This confronts us with the necessity for a model which does away with such distinctions. Instead we have postulated the smart city as an *integrated domain*. Under this view the digital and the human, the individual and the group, all hold equal status as nodes within a complex, dynamic autopoietic system known as a smart city. The concept of the smart city as an integrated domain provides a framework through which it is possible to consider issues which transcend traditional IT boundaries, issues deriving from complex interactions of multiple agents of multiple types. It provides a unified model by which to account for bidirectional interactions between management-level civic aims and individual device functions, between groups and individuals, between digital devices and humans. It further provides us with the ability to anticipate issues within mature smart cities without knowing the details of the technologies to come. Models like this, which seek to incorporate the human and the digital into integrated systems, offer our best hope of anticipating the future issues of smart cities.

- end of published paper -

7.2. Next Steps

The model presented above provides a framework for analysis necessary to answer the main research question, but does not address it directly. Within the terms of the above framework, the research question can be rephrased as "how can a digital environment restrict human autonomy?" The last stages of the above account introduce the integrated node as the operational unit within the integrated domain, but do not provide any detailed information about it. As the site of human existence within an integrated domain, and as the place where technology and the human interact, threats to autonomy will occur within the integrated node. Exploring the mechanisms by which autonomy can be restricted within the integrated node is therefore the next step. The following chapter offers an analysis of the mechanisms by which autonomy can be restricted through the integrated node and directly answers the main research question.

8. Autonomy within the Integrated Node

The book chapter presented here offers a second answer to the primary research question:

> "What are the possible threats to human autonomy generated by emerging ICT's?"

The chapter describes the mechanisms by which autonomy can be restricted within each of the integrated node's phases of input, processing and output. The aim is to provide a mechanism for considering ethical issues within future digital environments without needing details regarding the technology in question or the type of action being undertaken. The intent in so doing is to avoid predicating concerns on specific predictions of future technologies or specific human usages of them. This necessarily requires more detailed analysis of each phase within the integrated node, as well as supporting concepts.

The following section contains the book chapter as accepted for final publication. Minor changes in formatting and citation styles have been made to ensure consistency with the rest of this thesis.

8.1. Binding the Smart City Human-Digital System with Communicative Processes

In *Technology and the City*, edited by Michael Nagenborg, in-press.

This chapter offers a new way of viewing the complex phenomena which is humans living in a technologically advanced city. Using Actor Network Theory (Latour 2005), the concepts of autopoiesis (Luhmann 1995), symbolic capital (Bourdieu 1986), affordances (Hutchby 2001) and fields of contention (Bourdieu 1990), a new framework will be developed which provides a way of describing people and technology as a single unified system. We will use this framework to characterise the system's nodes and patterns of connection, identify fields of contention and how they impact technology design and derive consequent ethical issues.

The sophistication of modern technology is what separates modern cities from those of the past. Today's cities are becoming dependent upon advanced digital systems embedded throughout them. The life of a citizen in such a city is incomprehensible without understanding the technology of their built environment. We may therefore regard modern cities as complex socio-technical systems (Taylor 2005). The future for technology in cities is held in the image of the smart city, in which intelligent digital systems and personalised services dominate urban existence. If we can develop an all-encompassing framework to describe the smart city, we can use it to understand trends in urban technology today. By anticipating the ethical concerns in future smart city technology, we can identify current trends which may not be of concern today, but will give rise to issues in the future. Such knowledge offers the chance to change innovation paths now in order to prevent future issues developing in the first place.

The key characteristic of the smart city is the ubiquity of digital devices and systems. Smart cities are based on the foundation of an omnipresent sensor network which provides information about the inhabitants and environment (Balakrishna 2012). These sensors will communicate with devices and systems to provide analysis and response. Some of these communication patterns can be predesigned, but some will appear and disappear as people move about (*ad hoc networking*) (Guo et al. 2013). Digital devices will be embedded in the civic environment, the home and other personal spaces (such as the car), worn on the person and implanted within the body (Perera et al. 2014). While some devices will be embedded in traditional materials, the rise of smart materials will often render the distinction between device and material meaningless (Addington and Schodek 2005). Data will be processed locally (*fog computing*) as well as in the cloud (Bonomi et al. 2012). People will be tracked individually and in aggregate. The urban environment will change in response to the movements and needs of individuals and groups.

Ethical exploration of technologies is usually limited to issues associated with specific technologies, such as cars (Jaisingh, El-Khatib, and Akalu 2016) or location detection (Martínez-Ballesté, Pérez-Martínez, and Solanas 2013). However, many ethical issues emerge from the interaction of different technologies. Furthermore, smart cities will not be like today's cities with some digital technology laid over the top, any more than the modern city is just a Victorian one with cars instead of horses. The mature smart city will be a fundamentally different type of environment from any we have previously seen (Batty et al. 2012). Through all of history, the human built environment has been a largely dumb. It has not had the ability to initiate action, nor has it had the capacity to know us. The ability of the environment to obtain data and to respond to human action represents a fundamentally new type of built environment for human living (Picon 2015). The number, ubiquity, heterogeneity and invisibility of digital systems in the urban fabric will force humans to deal with intelligent spaces, rather than individual devices (Rolim et al. 2016). We must therefore cease to talk in terms of technical components and instead develop a vision of the smart city which encapsulates all aspects of its operation. The starting point for analysis is to identify the functional characteristics which can unite the various heterogeneous technical forms found in the smart city. These are ambient intelligence, artificial intelligence and robotics.

Ambient intelligence is technology which embeds input, processing or response throughout the environment (Ikonen et al. 2010). As a lived experience, people will not experience the smart city as individual components and discrete processes isolated within atomised sections of their lives. Instead, the smart city will be experienced as a seamless experience as one moves from house to car to work, from one context to another (Dainow 2017a). Similarly, smart cities will integrate data about people from multiple sources, creating a *digital environment*, a pervasive digital ecosystem which saturates the built environment, interacts with devices carried or worn by people and embedded within their bodies (Balakrishna 2012) to create an integrated sensing and response environment (Psyllidis 2015), an "intelligent information infrastructure" (ITU-T 2008, 2), or *ambient intelligence.*

Central to the vision of the smart city is algorithmic intelligence, or *soft AI* (e.g.: expert systems and software agents) (Komninos 2006). In addition to expectations that soft AI will provide core functionality through cloud computing, it has been suggested soft AI will need to be embedded at local level to support ad hoc networking (Komninos 2006), to provide contextual awareness (Augusto, Nakashima, and Aghajan 2010), and simply to handle the sheer amount of data (Komninos, Schaffers, and Pallot 2011). As a result, the smart city will be filled with many AI systems. These systems will need to interact with each other and the human inhabitants in myriad ways. These AI's will constitute a complementary, but distinct, form of consciousness to the humans within the smart city.

'Robotics' refers to ICT devices possessing the ability to move autonomously (Ikonen et al. 2010). Self-driving cars and drones are expected to exist within smart cities (Jaisingh, El-Khatib, and Akalu 2016). Much building construction will occur robotically (Tibbits and Cheung 2012). The home will see mobile IoT devices, such as robot vacuum cleaners (Cirillo, Karlsson, and Saffiotti 2012). Public spaces, such as a hospital ward or shopping mall, will see many different types of robot moving about (Mastrogiovanni, Sgorbissa, and Zaccaria 2010). The specific details of which robotic devices will exist within the smart city need not concern us here. The important point is that the smart city environment will involve both humans and machines moving within the same space. Just as the introduction of the car eventually led to the rise of pedestrian crossings, so an environment of moving machines will require adjustments in human behaviour to take them into account. Similarly, just as the car required the development of petrol stations, so the needs of the smart city's robotic inhabitants must inevitably require the development of specific urban structures and support systems.

The arrival of an intelligent, responsive environment containing autonomously mobile agents brings changes which must affect people. In this respect, the smart city will be created by its digital systems as much as by its human inhabitants. Device activity will respond to human activity. Humans will respond to device activity. Digital devices will have their own needs, many of which will require consideration by humans. This will sometimes bring the needs of the digital agent into conflict with the needs of the individual, just as cars require roads which gives them effective rights over people in some circumstances. Accordingly we must view digital devices as causal agents in a smart city on the same level as humans. If we are to develop a comprehensive view of the smart city, we must therefore include humans and digital systems as equals. Both are active agents capable of producing changes in the other.

Systems theory, long dominant in urban planning (Taylor 2005), permeates approaches to the smart city (Söderström, Paasche, and Klauser 2014), treating it as a complex system of systems (Bicocchi, Leonardi, and Zambonelii 2015). Actor-network theory (ANT) (Latour 2005) offers a systems approach which puts the smart city's digital agents into an interactive relationship with humans by regarding both as nodes in the same system. By treating both devices and humans as the same type of node within the same system, ANT provides a way of describing the causal effects digital agents have on humans and vice versa (Tabak 2015). A further advantage of ANT is the lack of need to specify the details of each node within the network (Law 1992). Instead, under ANT, agents may be a single digital device, an individual person, a group of people, a group of devices, or a mixture of these. However, ANT alone is not sufficient. The range and number of devices, their mobility, the deep interaction with humans, and the dynamic patterns of interaction generated as a result, means smart cities cannot be centrally managed, but must be self-organising, or *autopoietic*.

The concept of autopoiesis originated in biology to account for living organisms as self-sustaining systems (Varela, Maturana, and Uribe 1981). Systems theory holds that the defining characteristics of a system are not determined by the properties of the components, but by the patterns of their relationships (Boulding 1956). Autopoietic systems are those which can maintain or regenerate these patterns in the face of changing circumstances (Maturana, 1981). Niklas Luhmann introduced the concept of autopoiesis to the social sciences by treating society as a form of autopoietic system. Under Luhmann, the atoms from which social systems are built were *communicative events* between people. Luhmann defined communicative events as being composed of a sequence of information, understanding, and utterance (Luhmann 1986b).

This communicative atom is also visible in digital devices and systems. It is the minimum necessary characteristic of any digital device that it be able to accept input ("information"), process it ("understanding") and produce output ("utterance"). All digital systems are formed from groups of digital devices acting in such a fashion. However, smart cities will not be constituted by systems in which digital devices communicate only with each other. Humans and digital devices will engage in communicative patterns across the human-digital divide, as well as with each other. Furthermore, significant portions of human-to-human communicative activity will be mediated by the digital environment, while that digital environment's communicative patterns will respond in myriad ways to every human action. Under such circumstances, few nodes can be said to be purely human or digital; the overall patterns of communicative flow will be the result of both human and digital actants - an *integrated domain.*

It is important to bear in mind ANT is not explanatory. ANT is more method than theory, and the appellation of 'theory' has been declared misleading (Latour 2005). The concept of the integrated domain is not intended to explain relationships or trace cause and effect. ANT provides a system of describing phenomena which cannot be attributed to exclusively digital or human causes, but does not seek to explain why or how. The descriptive framework offered here is intended to provide a language for such explanations.

8.1.1. Integrated domains

An integrated domain is an autopoietic socio-technical system comprised of two collectives integrated into a mutually dependant partnership. One collective consists of a human *smart society*. A smart society is a human society which exists within an ambient digital environment, such that human intersubjectivity is mediated by digital technology (Hartswood et al. 2014). The other (non-human) collective is a system of digital devices and networks. Operationally, neither collective possesses strict boundaries against the other. Instead, they intermingle to such a degree it is impossible to account for phenomena within the integrated domain without reference to both. An example is the search engine manipulation effect, which has shown that the order in which a search engine lists items strongly influences the chance of these items being read and being believed (Epstein and Robertson 2015). Search engines are significant cognitive tools for many people, which analyse people's behaviour in order to produce search results personalised to them (Dillahunt, Brooks, and Gulati 2015). Thus, the user's knowledge of the world is mediated by logic created by their behaviour in combination with the search engine's algorithms. The user's knowledge of the world no longer derives from their own efforts, but from the inseparable fusion of human and digital. When the human and the digital are inseparable, we have entered the integrated domain.

The node which binds the integrated domain as a system is the *integrated node*, a unit of operation consisting of input, processing and output. In ANT terminology, an integrated node is a mediator, a node in a system which transforms what flows through it (Latour 2005). A integrated node may be a person, a single device, a group of devices, a group of people, or any combination thereof. A integrated node is not defined by what it is made of, but by its mode of operation.

As the digital develops more cognitive capabilities through AI, so its impact on the human becomes more varied and more complex. The internal nature and logic of these digital systems, their *logos*, therefore becomes more important in understanding their outcomes. As a consequence, it becomes imperative that we consider the digital perspective - how the world looks and is understood by digital systems (Picon 2015). The terms 'looks' and 'understood' do not indicate anthropomorphism. Digital cognition is not the same as human cognition. However, if cognition is accepted as "the action or faculty of knowing taken in its widest sense" ('Cognition, n.' 2018), then it is clear that digital systems do possess a form of cognition. The most important aspects of the digital perspective are time, design, logic, epistemology and energy.

Descriptions of digital systems are inherently time-bound. Digital systems only "exist" to the degree that data flows through them. Without data on which to operate, a digital device is inert, no more able to cause effects in the world than a rock. Accordingly, a digital system is not best considered as a state but as a process (Dodge and Kitchin 2005). For example, a service commences with anticipation of user requirements, organises processes to deliver the service to the user, is processed by the user (possibly mediated by other devices) and terminates in a stream of output data comprised of user activity and the recording thereof. Integrated nodes are regions through which data and activity flow, defined by the manner in which they change what flows through them. Digital systems cannot be understood as actants by considering their material composition, only through understanding how they operate over time.

Digital constructionism is the view that there is nothing natural, and little inevitable, in the manner any digital system accomplishes its task. For example, it is often assumed that improving online security must require reducing online privacy, but these two values are not in competition. Enhancing privacy can actually make systems more secure (Langheinrich 2001). The assumption that security and privacy are oppositional reflects current practice, not the necessary characteristics which must adhere in any and all digital systems (Dainow 2015b). Any system which improves one by reducing the other is a result of decisions made by developers, who chose to code in such a manner. The algorithms, interfaces and other implementations of digital design are therefore to be understood as artificial and engineered social constructions, and always open to alternatives (Dong 2004).

All digital systems stem from the place of tool making in the human. It is the primary defining characteristic of the digital that digital systems are created by humans for specific purposes. We do not create any digital systems simply so that they can exist, returning nothing to us. All digital systems are therefore inherently teleological systems. Any attempt to understand them must consider the purpose for which they exist. We can only arrive at a complete understanding of other goals, processes and values within a digital system by considering them in reference to the ultimate purpose for which that system was built. We must therefore consider digital systems in terms of who their primary beneficiaries are and the benefits they derive. Processes within digital systems are driven by their *digital logos*, the heuristics, values and other elements which govern the design of digital cognitive processes and thus determine their outcomes. The nature of the digital logos is therefore important to those who desire control. Consequently, contention and power dynamics are concentrated around the digital logos in the form of competition for dominating values and heuristics (Albrechtslund 2007). The overarching function of the digital logos is to serve the system's ultimate purpose. The ultimate purpose may be different from the obvious purpose. The obvious purpose is the reason why users make use of the system. For example, a commercial navigation system's obvious purpose is to guide a user to their destination. However, as a commercial system, its ultimate purpose is to generate profit. During design, elements of the system will therefore be evaluated against their profit-making capacity as well as their capacity to aid navigation. Potential elements which help the user but compromise profit generation will be unwelcome, and may not be included (Bogliacino and Pianta 2013). The user may not get the best solution if such a solution would reduce profitability. All systems, especially those operated for profit, must therefore be regarded as, at least potentially, compromising the user in favour of some other goal. We can never assume that the user is always the primary beneficiary of any digital service.

By its very nature a digital device cannot know the physical world. What a device knows is only what its sensors generate as input. For example, an air conditioning control system does not respond to actual room temperature, but to what the sensors tell it the temperature is. If the sensor gives a false reading, the air conditioning will still respond to what the sensor tells it, not what the reality is. Hence digital devices do not know the physical world, but inhabit a totally digital environment. What they know of us is determined by our representation in digital data. Smart city systems will not respond to us; they will respond to our digital representations. If those digital representations are incorrect or incomplete, then the analysis of us and our needs will be compromised (Mittelstadt et al. 2016). *Digital scepticism* is the epistemological position that digital systems can only detect that limited set of reality which they are designed to receive as input. If a system is not designed to process a certain input, then the corresponding aspect of the world does not exist for the digital system. Should this aspect of the world represent an important goal to the user, their attempts to achieve that goal will be frustrated. When combined with digital constructionism and the politics of the digital logos, the digital system's knowledge of the world becomes a product of stakeholder interests, not some "objective" view of reality.

Every process uses energy. This means we can validly apply many concepts related to energy to both digital processes and human interactions with digital systems. We can talk in terms of capacity, frequency, resistance and so forth. We can apply these same concepts to the human as well as to the digital because we do not depend upon any particular material composition for the concept of energy to be applicable. Consequently, we can discuss both digital and human systems and the interactions between them in the same terms. The amount of energy required to utilise a digital service is an important factor in its usability. Energy is expended for a purpose, as a cost towards achieving a goal. Goals have a cost threshold, beyond which the cost is not worth the goal. Increasing ease of use is a way of reducing the energy required for a task. This is a significant factor in the spread of digital technology and an important driver of innovation (Fontana and Nesta 2009). By contrast, *nudging* is a process of gradually leeching energy through micro-thresholds in order to produce a strong cumulative effect, either to discourage input, prevent or influence processing, or reduce output (E. J. Johnson et al. 2012).

8.1.2. Contention and affordances

As creations of human society and as sources of value, digital systems are inevitably included in the *fields of contention* (Bourdieu 1990) which permeate other aspects of society (Pursell 2007). Many of the characteristics of any digital system can be sites of contention (Mansell 2017). Any aspect of a digital system can be designed to increase domination more than to serve the user. Characteristics of digital systems should therefore be considered as potential strategies for competition. Within the Integrated Domain digital actants are not passive spaces, but support or inhibit the competitive strategies of human actants. Bourdieu's theory of multiple forms of *capital* (Bourdieu 1986) allows us to understand human-digital interaction as being engaged in generation, exchange and transformation of capital. Digital systems can produce multiple forms of capital – economic (OECD 2013b), cultural (Shifman and Nissenbaum 2017), social (Jiang and Carroll 2009) and political (Epstein and Robertson 2015).

Generation of non-economic capital in digital systems occurs through *affordances*, which bring cognitive associations to sensory input. Any technology may be understood in different ways, according to a person's education, social environment and other factors (Hutchby 2001), all of which are constrained by the capabilities of the technology in question (Norman 2002). These dictate a person's affordances; how they perceive a technology and how they understand what they perceive. The concept of affordances holds that perception does not consist of a physical activity onto which understanding is overlaid *post hoc*. Instead, the perceptive process itself contains cognitive elements, such as motivation and context (Gibson 1986). We use affordances to build conceptual models of how something works (Hutchby 2001). There are always two conceptual models involved; the model the designers held when they constructed the artefact and the user's model when they use it. These models are never the same and user error or difficulty is usually the result. There is a significant body of analysis which argues that the design and operation of digital services today is primarily tuned to the exploitation of users to their disadvantage (Vaidhyanathan 2012; Andrejevic 2012). Owners use marketing to manipulate values so as to attribute symbolic capital to use of their systems (Bernays 1928), such as creating socially aspirational personalities whose use of a system makes it desirable to their followers (Lamb and Kling 2003). Designers of digital systems use control of development to determine which forms of capital are generated through use of their systems (Dainow 2015a) and which are repressed (Cirucci 2015). A common example is making privacy controls hard to locate in order to stop people using them (Liu et al. 2011).

8.1.3. The Integrated Node

As a system, we do not understand the smart city's digital environment as being composed of states so much as processes. A *integrated node* is a node in an Integrated Domain which binds the system together through the activities of input, processing and output. This is the model for mediators of any composition within the Integrated Domain, whether composed of single or multiple actants, digital or human, or any combination thereof. Integrated nodes are recursive. They are usually made from autopoietic systems of sub-nodes, each of which is an integrated node itself. Each of these will usually be constituted by their own autopoietic system of integrated nodes, and so on. For example, a person may interact with their digital environment through a combination of their physical body and carried digital "assistants" such as a smartphone. Here the one integrated node is composed of both the human and the digital devices, both of which can individually accept input, process it and produce output, and thus are nodes in their own right. However, to be an "integrated" node, the node must be composed of both human and digital components. However, the triphasic pattern of input, processing and output is the atomic pattern within all nodes, whether human, digital or integrated. Hence, the integrated node constitutes both the atomic unit and the essential pattern of interactions between these atomic units, transcending normally distinct ontological boundaries.

Many integrated nodes will be concerned with internal states of operation within both digital assets and humans. Much of the communication between digital agents concerns management of digital technology. This monitoring represents a form of digital self-awareness, though 'self-awareness' does not imply sentience. Meanwhile, the human element confers on the Integrated Domain the human level of understanding and meaning-giving. An Integrated Domain is therefore a self-aware system which two forms of awareness – human and digital. We are already witnessing early examples of hybrid systems which combine human and digital devices. *Social machines* are "socio-technical systems which involve the participation of human individuals and technological components…able to extend the reach of both human and machine intelligence [by] supporting capabilities that less integrated systems might find difficult to accomplish" (Smart, Simperl, & Shadbolt, 2014, pp. 55–56). Social machines are early fore-runners of the Integrated Domain's integrated nodes which will bind together the fabric of future smart cities.

The integrated node is both its own field of contention and a site within other fields of contention, in which attributions of symbolic capital determine value, some of which are convertible into economic capital. The initial basis for contention is the degree to which the user's needs align with the service provider's or designer's aims. Where the user is paying directly for the service and has a choice of service provider, the interest of the service provider is primarily to maintain the satisfaction of the user. However, where the service is free to users but paid for by other means, the service provider's concern for the user is merely to keep them using the service while value is derived from that usage of the system by other means.

Each person inhabits a digital environment comprised of a variety of devices and systems. Some of those devices are "personal" in the sense that they are worn, carried, or embedded in the body. In addition to providing direct services to the user, these devices mediate between the person and their digital environment. The smart city will not interact with people directly, but with this assemblage of human-with-personal-devices. At the level of the Integrated Domain, each individual is thus located within an *Integrated Personage* comprised of themselves plus their personal digital devices. Should one subscribe to theories of the extended self (Belk 2013), an Integrated Personage may be regarded as a self which has been extended into the personal digital devices. However, the concept of the Integrated Personage is not dependent on the concept of an extended self, but on practical necessity - many digital systems are not designed to interact with humans, only with other digital technology, and so humans must use digital tech to access these systems. *Integrated Devices* are those devices used within the Integrated Personage. They have a special relation to the human by virtue of their role as critical mediators with the Integrated Domain and within the human smart society. They may also have, under some accounts, special relation to the user as extensions of the self, especially if they are embedded within the physical body. As a result they have a special ethical status not accorded to other owned objects.

8.1.4. Ethics and the Three Modes of the Integrated Node

We need an account of the technological city which does not function at the level of the thousands of individual technical systems of which it is made. Not only is such a level of analysis is too complex, it is misleading because technologies do not work in isolation but are experienced as a unified digital environment. Likewise, we need to avoid separating the digital from the human because it is often impossible to understand the one without reference to the other. In similar fashion, consideration of individual ethical concerns is to be avoided. Individual ethical concerns are many and frequently specific to particular technologies, usages or historical circumstance. The approach taken here is to instead identify a single point of ethical concern which is not dependent on specific details. This single point can be used to identify features of ethical concern which could be found in any digital environment. Autonomy is just such a foundational ethical value. It forms the legal and ethical basis for human rights and provides the justification for treating people as we do under international law and democratic politics (Dworkin 1988). All rights given under the Universal Declaration of Human Rights derive from the premise that all humans have the capacity for autonomy (Marshall 2008). The capacity for autonomy is what gives humans dignity and the right to be treated as an end in themselves, not a means to an end (Kant 1998). Autonomy is thus the single point of ethical interest with the widest range of implications and shall therefore be our point of ethical concern.

However, there are many different definitions of autonomy (Dainow 2017b). The term was initially restricted to morality by Kant, who argued that all humans had the innate capacity to determine what was morally right and wrong (Kant 1998). This version of autonomy has become known as 'moral autonomy.' Since Kant we have added political autonomy (participation in political processes) and individual, or personal, autonomy (being the final determinant of one's actions) (Christman 2015). In addition to these spheres of application, there are many different accounts of what constitutes autonomy in any sphere. Here the focus lies on the internal processes by which one exercises autonomy. The general concerns are to distinguish the degree of external influence which can be allowed before we can say autonomy has been lost and what, if any, specific cognitive elements are required in order for self-reflection to count as autonomous (Dainow 2017b). Even the same definition of autonomy can lead to people living incompatible lives because they hold different values. However, it is possible to work with a generic understanding of autonomy without getting bogged down in these details by allowing that it is for each person to determine how they want to define autonomy for themselves. This means we need to assess digital environments in terms of the degree to which they enable or prevent the exercise of all forms of autonomy. At the level of the Integrated Domain, the Integrated Personage occupies the point of autonomy for the individual and so our account of the operations of the integrated node shall focus on the ways in which autonomy is in play within the input, processing and output phases of an Integrated Personage.

8.1.4.1. Input

Input may be defined as change at the interface between the integrated node and its environment. It is through design of digital interfaces that the most significant efforts at control occur within the Input phase. Interfaces are material elements which generate energy for internal processing. Data is a cognitive form of energy and may be analogue or digital. When meaning is applied to data or use made of it, data becomes information (Ackoff 1989). Different meanings may be applied to the same data, just as different uses can be made of it. The same data is thus capable of giving rise to variable amounts of information (Floridi 2008). Someone's capacity to control their input is both empowered and constrained by the nature of the interfaces which exist between the human user and the digital environment. Control of interface design and operation is therefore fundamental to domination and commodification (or "monetisation") of the digital environment.

One cannot process something which has not entered the system. Input therefore determines the constraints of processing. Control of input is a fundamental determinant of the exercise of autonomy. No matter how one defines it, autonomy is always dependant on one's knowledge of the external world because it is with this knowledge that one makes deliberations. One important way in which input is restricted is through *epistemic control.* Epistemic control is the strategy of restricting the delivery of data or designing its delivery so as to hide information. A common example is the creation of privacy policies which are lengthy and difficult to understand. By doing so, the privacy policy becomes too large or too complex to be accepted as input by the user (McDonald and Cranor 2008). Because the policy cannot be input, the website's privacy practices cannot be processed, and so the user cannot generate a response (output) which the service provider wishes to avoid (such as not using the service) (Acar et al. 2015). Epistemic control is also exercised by simply hiding what is being done. For example, digital systems may make use of hidden technologies, such as Flash Long Storage Objects, which bypass privacy systems (Soltani et al. 2010). User systems which attempt to preserve privacy by blocking cookies are an example of *epistemic contention*, in which two parties attempt to gain epistemic control of each other. In effect, the user's cookie-blocking technology is an attempt by the user to limit the tracking organisation's input, and so prevent the tracking organisation processing information about them, while the tracking organisation's attempt to hide their tracking is an attempt to prevent information about their practices becoming input to the user. This exposes an on-going conflict between service provider and user, in which each seeks epistemic control over the input of the other as a means of influencing their respective processing. A more subtle strategy for epistemic control lies through association and design, as we have seen with the search engine effect (Epstein and Robertson 2015), whereby the mode of presentation of data at the interface can influence the affordances through which it is input to the human. More complex factors emerge in concerns as systems become more sophisticated. Factors such as cultural bias may mean the input systems for ambient intelligence, affective systems (which seek to detect emotion) and voice recognition, force people into behaviours which are foreign to their culture in order to use the system (Soraker and Brey 2007). Epistemic control can also be achieved through preventing restriction of input, effectively "force feeding" data into the integrated node. Advertising technology which bypasses ad blocking is an example of such force feeding, as is subliminal messaging (Verwijmeren et al. 2011). The actions of data brokers in seeking to keep their very existence hidden (Federal Trade Commission 2014) are an extreme example of epistemic control.

The designer of a digital system may restrict input intentionally or through ignorance. Where this is intentional, it may be done either to help the user, make life easier for the developer, or exploit the user. Where it is done to exploit the user, it constitutes an intentional attack upon their autonomy. Many theories of autonomy hold that autonomy is preserved if the person would have consented had they been given the opportunity to do so at the time (Dainow 2017b), so hidden factors do not necessarily constitute a restriction on autonomy. However, preventing choice *post hoc* does constitute a restriction of the user's autonomy no matter what the action. Where this prevents someone taking actions they would have chosen if they had possessed full information, it also constitutes a restriction on their freedom. Where input is restricted to suit the service provider it constitutes a secondary, but still intentional, attack on the user's autonomy, in that the user is not being considered as an end in themselves, but as a means to the service provider's end. Where input is restricted to help the user, it can be described as an accidental restriction on autonomy and should be weighed against the benefit derived. This happens fairly frequently. For example, it is often easier for users to restrict input by forcing people to choose from pre-set options, rather than allowing them to type freely. A free choice opens the possibility of unprocessable input, and so requires checking which may force the user to retry. As a result, restricting users to pre-set choices is the preferred strategy in design.

8.1.4.2. Processing

Processing is the bridge between input and output. Processing therefore heavily influences which connections are made at output, although output is also constrained by material limitations. Input is what is processed. We use the term 'energy' for that which is input, including both data and physical activity (such as hearing aids and traffic lights). Because affordances append attributions of meaning and implied use, processing is strongly influenced by the affordances with which the input was received.

Processing may be described as the reorganisation of input energy patterns through manipulation. The heuristics of this manipulation are driven teleologically by the goals for which they are deployed. How input is transformed is also determined by the material composition of the integrated node and its internal processing structure. The processing structure may be divided into a material component and a cognitive component. Examples of the material component include biology and chip design. The material component has a determining effect on many aspects of the data, such as granularity, processing speed, computational complexity and so forth. The factors affecting processing are always defined so that they can be applied to a device, human, individual or group, no matter whether digital, organic or both. Since it is attributions of value which determine the heuristics of processing, contention occurs for attributions of value in an attempt to influence the processing. For example, placing values on using digital devices in certain ways (Zhao, Grasmuck, and Martin 2008) is an attempt to influence the way a person processes information using these devices. In an Integrated Personage processing is also instantiated in communication patterns between the Integrated Devices, and between these devices and the human. For example, a GPS device communicates location information to a mapping device which then communicates instructions to a bluetooth headphone which then communicates sound to the human, who then changes direction and thereby communicates new location data to the GPS device. It is not only the data processed and transmitted which constitutes the processing, but the devices themselves as Triphasic Nodes within the autopoietic system which is the Integrated Personage.

Many of our existing issues with regard to ICT ethics today are focused on processing; issues of algorithmic justice, the moral status of autonomous systems, and use of personal data, are all examples of concerns over processing. One of the most important ethical issues here is the role of previous experience. It is indubitable that individuals bring their previous experience to their processing. Physical practice to develop muscle memory and trained reflex, central to music and sports, are a way of encoding previous experience as processing structures in the physical body. Education is a way of producing previous experience which can be used in cognitive processing. Previous experience is held to be a critical factor in processing by designers of personalisation systems, which analyse our history in order to predict and influence our future actions (Schneier 2015). The internet economy today is based on knowledge of previous experience (OECD 2013b), as are all personalisation services of the future.

Because ANT is not an explanatory theory, it brackets how processing works. Being silent as to the internal details of processing, the exact mechanism by which human processing occurs may be described according to one's theories regarding the nature of the human mind. All that is required here is an acceptance that people have internal states of some form, determine their own actions in some way, and respond to changes in their environment. For our purposes, it is sufficient to say that processing involves the production of information and determination of action, and that input comes with a use value (Hartson 2003), which implies how it is to be processed.

It is important to bear in mind that each human exists within the Integrated Domain at the level of the Integrated Personage, an autopoietic system composed of human + digital. The Integrated Personage is therefore the point of autonomy within the Integrated Domain, not the individual. Since autonomy involves self-governance (Kant 1998), the ability to have control over processing within the Integrated Personage is necessary for autonomy in an Integrated Domain. The individual must be able to control processing to the degree that they are able to generate output intended to produce the states of affairs that the individual desires. It is not required that they are actually *able* to produce their desired states of affairs. Autonomy is not threatened simply because the world does not give in to whatever you wish. But the individual must be able to *try* (Christman 2004).

A person may be unable to exercise autonomy in processing due to constraints created by designers. *Internal contention* is contention for the manner in which the Integrated Personage processes information. *External contention* is competition as to whether processing occurs within the Integrated Personage or is done by an external integrated node. Internal contention occurs through attempts to influence how the Integrated Personage conducts its processing. Simple examples include blocking interoperability with a device from a competing manufacturer, controlling what apps can be installed on a device, or restricting devices from accessing certain files or formats. A common technique in many systems is to discourage processing by increasing the energy required, for example the use of obscure legal terms in privacy policies. One has no difficulty physically inputting the words into one's mind. However, where the terms are not understood, the affordances providing semantic understanding at input are absent. In order to process these terms, one must look them up in a dictionary first. This raises the energy required to process the term significantly, and may reach a level which exceeds user's willingness to continue (McDonald and Cranor 2008). It has been argued that is exactly the purpose behind using such terms (Schneier 2015). Such design elements are examples of the way in which people can be nudged into changing the manner of their processing through design (Acquisti 2009). When an operating system provider embeds their web browser into the interface in a manner which makes it hard to avoid using, as Microsoft once did (Fisher 2001), they are similarly seeking to take control of an element of the processing within one's Integrated Personage.

External contention is contention over whether processing will occur within the Integrated Personage or within an external integrated node, such as a cloud service. Systems which upload data to the cloud for processing are transferring cognitive load from the Integrated Personage to the service provider's integrated node. Systems of rentier capitalism for Integrated Devices, such as software-as-a-service, represent an attempt to own portions of the individual's Integrated Personage. If books and similar digital services represent offloading of personal cognitive capabilities into digital devices, then statements of ownership over content, such as we see with Kindle (Amazon 2017), constitute claims of ownership over the individual's cognitive capabilities. The argument from tradition that digital devices are optional tools is contrary to modern circumstances. The right to hack and to repair one's own devices contends with intellectual property initiatives and rentier capitalism for control of how individuals process within their own Integrated Personage. A digital service or device which is so commonly used that not having access to it impacts one's ability to participate in society constitutes a component of the smart society. Lack of access means one cannot fully participate in the smart society. As technology spreads deeper into the built environment, so it comes to assume certain minimum technological capabilities on the part of the inhabitants (Dodge and Kitchin 2005). Increasingly, the city's digital environment has expectations that the Integrated Personage possesses certain capabilities - certain input and output channels and certain processing capabilities. Where an important aspect of the Integrated Domain becomes dependent on such capability, that capability becomes necessary for a full life and full participation in the smart society. At that point one acquires, as a citizen, a right to that capability. To the degree that this capability is necessary for the living of a full life, it becomes a human right.

8.1.4.3. Output

Output connects an integrated node with other integrated nodes. One integrated node's output is the next integrated node's input. We can therefore talk in terms of an Output-Input channel (or *O-I*). For the purposes of this discussion, output may be defined as energy released from an integrated node in the form of physical movement or data. Some output is intentional, though cases of hidden surveillance show that output need not be originated by the user, or even known to them (Sandoval 2012; Federal Trade Commission 2014). Output also occurs as a by-product of existence because all digital processes leave a *data shadow* (Westin 1970) in the form of temporary files and logs.

Because the function of output within the Integrated Domain is to connect with other integrated nodes, output must be *coherent* with its environment. Here 'coherence' means compliance with shared standards, such as languages, protocols and data formats. The requirement for coherence grants power to those who can control it. A significant portion of the contention within digital technology is for control over standards because it confers domination over systems dependent on those standards (Cusumano 2010). Control of connectivity accords one the status of a gatekeeper, able to control who uses an O-I channel and what they can send through it. The gatekeeper role in connectivity is one of the major points of financial return in digital services. Websites act as gatekeepers for connecting advertisers to users and extract value therefrom, charging an entrance fee for a presence in the user's input stream (Deighton and Kornfield 2012). Telecoms companies generate income from connecting other digital systems to users (Huston 1999). Debates over net neutrality show that many believe service providers can use this position as gatekeepers to extract economic capital (Pil Choi and Kim 2010).

As techniques such as object-oriented programming (Rumbaugh et al. 1991) and Digital Object Architecture (York 2016) demonstrate, internal processing need not use the same formats as are required for output. As a result, output often requires a translation layer to convert internal information into formats which can be accepted as input by external integrated nodes. For output to humans, this may mean displaying information in a readable format, with consequent constraints on minimum text size, display time, using a particular human language, and conforming to graphical conventions in digital design (McKay et al. 1997). Coherence means providing digital devices with data in a format they can accept, hence the need to create keyboards for human output intended for digital systems. The Graphical User Interface (GUI) and similar systems, such as HTML, provide formats for both data and affordances (Galitz 2007). API's, networking protocols and similar systems provide shared output formats which can bind digital devices into communicative systems. Physical output is seen in the activation of motors and switches, and the movement of muscles and limbs. Much of the intelligent functionality of the smart city is location-dependent (Picon 2015), such as controlling traffic flow or delivering location-based services to inhabitants. The movement of the Integrated Personage within the digital environment will be accompanied by a constant flow of output based around movement through an urban space whose digital characteristics change as fast as its physical characteristics (Dodge and Kitchin 2005). The Integrated Personage will manifest within the digital perspective as a moving cloud of output.

Output currently constitutes the primary point of value extraction for digital service providers where the recording of user behaviour is monetised through data mining (O'Neil 2017). As a result, commercial surveillance is one of the foremost ethical concerns regarding output. The ethical status accorded to human beings on the basis of their possession of autonomy (Marshall 2008) confers special ethical status on data output by the Integrated Personage. This concern is greater where usage of output results in restrictions on autonomy or derivative rights. Autonomy is limited when people's digital systems do not allow them to emit the particular form of output they wish. For practical purposes, we cannot consider something a limitation on autonomy which is physically impossible. If it is not technically possible to instantiate my preferences, then my autonomy is not reduced. Nonetheless, in many cases, systems are designed to make outputting particular desires or in particular ways difficult or impossible - nudging or forcing people into using approximate variations which can be commoditised (Dainow 2015a; Cirucci 2015).

It is in the user's interest that their output is tuned to optimal delivery of the services they desire, in the manner they desire, under the conditions they desire. It is in the interest of a commercial service provider that they offer the services in the manner which generates the greatest income. These two interests are not necessarily aligned. Where they are not aligned the risk arises of reductions to autonomy through the use of that service. Where the service is non-essential we may say that autonomy is not threatened because one has the option of simply not taking the service. However, where the service is essential, or has achieved the status of a public utility (such that participation in society becomes difficult without it), then conditions under which one is forced to accept that service risk threatening autonomy. It is possible therefore, to treat certain aspects of service delivery as public utilities and impose limitations on allowable terms of service. It seems most in accord with the capitalist model to promote competition by requiring alternative service providers offering alternative terms of service. The construction of a smart city in which the citizen has access to alternative providers for each and every service at each and every location is feasible but would require an open technical infrastructure, similar to that offered by TCP/IP and HTML. Such open platforms enable any designer to create new systems in confidence their output will be coherent with the digital environment. No participant in the TCP/IP or

HTML ecosystems can prevent a competitor arising (no one can become a gatekeeper to the web). Achieving this situation in the smart city would require standardisation of many protocols, but we can have done this before with many other standards (Mattli 2001), and indeed it may be argued from history of utilities that it is almost inevitable that we will do so in the future (Clifton, Lanthier, and Schröter 2011).

8.1.5. Summary and Conclusions

We have passed the point where technology can be treated as a dumb hammer, whose use and impact on others is totally dependent upon the human who wields it. Society, now and into the future, can only be understood by considering the human and the technological in combination. This combination results in systems giving rise to emergent features not predictable from the components. We therefore need a theoretical framework, a language, specifically designed to handle that fusion as a thing-in-itself, not as an analogue of traditional concepts. Actor Network Theory, combined with the concept of autopoiesis, offers the vision of an Integrated Domain in which the mediating nodes which provide the motive power and logic to the autopoiesis of the system are created of the interactions of humans with digital systems. The human and the built environment of the city do not directly encounter each other but interact through the medium of digital systems. The nature of their interaction is such that it cannot be split into independent components of the digital and the human, but combines them to the degree that they are indistinguishable. In terms of effects, they become one cause. They also give rise to emergent properties not predictable from their individual characteristics and not present without both.

As a dynamic system, the Integrated Domain consists of the transmission of energy in various forms through integrated nodes, which are mediators constituted by a three-phase process of input, processing and output. These nodes are therefore, in terms of their operational characteristics, events and processes more than they are objects and states. The ontological priority of process derives from the fact that integrated nodes cannot be understood, and do not exist, unless they utilise processes over time. It is true that digital devices will have a material composition, just as humans do. However, the material constituents are separable and understandable without reference to the other. By definition, the Integrated Domain is the ontological level at which the digital and the human are fused. That fusing occurs only as a process. Therefore, by definition, material composition and static states are not part of the

Integrated Domain. It is only when nodes produce output and pass it as input to other nodes that the Integrated Domain comes into existence. Consequently, it is always through process and connection that we see the Integrated Domain.

Analysis of ethical issues and contention within the Integrated Domain must therefore focus on connectivity and processing, seeking to find mechanisms by which competing stakeholders pursue strategies for control and value extraction. We understand what is being transmitted through the system as energy, using the term 'energy' generically so as to allow for both material and cognitive (or symbolic or mental) transmission and activity. In empirical terms observed phenomena are almost always a combination of the two, just as they are a combination of the digital and the human

From an ethical point of view, and to the individual, the most important integrated node is their Integrated Personage. The Integrated Personage is the active field within the digital environment constituted by the human in an inseparable fusion with their personal digital devices, such that the aggregated output from the Integrated Personage is a unified product of both human and digital, including not just the individual operations and functions of the component devices, but also the patterns and content of their interaction. The Integrated Personage, like any other integrated node, is an element of the Integrated Domain and therefore a combined entity in which the human and the digital cannot be treated as distinct.

Input overlaps with output. The output from one integrated node is the input for the next one. This output-input (O/I) connection constitutes the systemic relationships between the nodes maintaining autopoiesis within the Integrated Domain. Output must be coherent with the digital environment in order to connect with it. This means that control over what constitutes coherence dictates control over how, and possibly what, can be output, and to whom. Examples include network protocols, programming languages and data formats. In the absence of contrary forces, we can expect contention over standards to continue as the smart city develops. Given the wide range of functions, spaces, contexts, volume and variety of digital devices and systems within the smart city, we can anticipate many more opportunities for such competition. On the basis of the history of digital innovation, we expect that multiple stakeholders will participate in this contention, few of whom can be assumed to have the user's best interests as their primary goal. In the absence of countervailing forces, we can therefore anticipate that the individual user of the service is unlikely to receive the best solution to the problem. Rather they are more likely to receive something just good enough to tolerate, while their use of the service is exploited, both in terms of the service delivery and the design of the systems. As the Integrated Domain rises to become more sophisticated and more the dominating feature of modern urban society, so the potential for domination by those who do not have the user at heart becomes more likely and potentially more damaging. We do not know what the future will be like, but by anticipating the flows within future urban socio-technical, we can identify where human autonomy will be threatened and how.

- end of published chapter –

9. Conclusions

9.1. Summary of findings

Each publication offers some mechanism by which ICT's can restrict autonomy. This section brings all these factors together.

9.1.1. Computer science mentality

Many who are happy with those problematic elements which restrict human autonomy do not believe alternatives are possible. For example, many believe that delivery of personalised services requires reductions in personal privacy. However, we have demonstrated how Langheinrich's Principle's for Privacy Aware Ubiquitous Computing and emerging alternative architectures offer many ways to deliver personalised services while also maintaining privacy. Others believe that the creator of an autonomous system has no ethical responsibility for the actions or effects of that system. Under this view, once an autonomous system is activated, human responsibility terminates. However, this position is based on a highly restricted and unusual understanding of the manner in which ethical responsibility transfers between causative agents. The position requires that services be treated as independent of any human cause in terms of design, environment, purpose, how they are used and who benefits. In fact, once autonomous services are understood as existing within a field of human practice, the ethical responsibility of the designers becomes self-evident. Those who design and deploy ICT systems therefore need to accept that they bear the ultimate ethical responsibility for the effects they generate. To put it bluntly, you cannot regard yourself as a good person if you build evil software. Finally, the client-server service architecture has been accepted reflexively by developers and users, despite evidence of more effective alternatives in many cases, and has been supportive of autonomy-reducing organisational and economic models. Alternatives are possible which do not enable problematic organisational or economic models to arise with such ease. The solution to these issues is to educate programmers and leaders that there are viable alternatives to customary practice and that they have the power, and a moral responsibility, to explore them.

9.1.2. Treating users as captured resources

The ability to express and act freely in digital environments is restricted by power structures designed around extracting money from user activities by spying on them and by designing systems which facilitate only those activities which support this surveillance and monetisation. In other words, users are not treated as the beneficiaries of the systems, but as commodities within a larger system, most of which is intentionally and vigorously concealed. User autonomy is further restricted when the data gathered about someone is used against them, either to manipulate their behaviour or to influence how others treat them. This system combines interface design, system architecture, business model and legal framework to alienate users from the products of their own digital activity. Steps are put in place to prevent user resistance by abandonment of the open standards which created the web, so that user activity is imprisoned within data silos owned by the service provider. These factors combine to alienate people from each other because they are forced to communicate through these exploitative social networks and by the chilling effect of ubiquitous surveillance. Finally, users are alienated from themselves and their own potential through the imposition of fetishizing affordances, such as the concept of the personal brand, through their limited control over their own digital data and through the use of personalization technologies which trap the user in a "filter bubble," restricting their ability to discover the unexpected, the unusual and the uncommodified. On this basis, users are being treated as a means to an end, not as an end in themselves. Such treatment is the very definition of a violation of human dignity and therefore a violation of user's human rights.

9.1.3. Surveillance

The presence of a surveillance system is a necessary precondition for the use of personal information by others and for personalisation services. Control of this data by the individual is therefore an effective way to retain autonomy. Thus control of privacy is an enabling right to autonomy. Autonomy-restricting systems depend on hidden, uncontrollable or unavoidable surveillance.

9.1.4. Insufficient user control

Restricting user control is a necessary precondition to restricting autonomy. Lack of control can be accomplished through lack of knowledge on the part of the individual as to what is occurring, through "take it or leave it" terms of use combined with a lack of alternatives and through the lack of suitable user configuration capabilities, or the concealment thereof. In some cases, not just functions, but entire industries are concealed.

9.1.5. Insufficient user configurability

Autonomy can be threatened by lack of variability in configuration options, such that the user's desires, style of operation or range of outcomes cannot be enabled. This lack of variation can stem from ignorance by the ICT developers, insufficient understanding of the user population, insufficient granularity of configuration options or systemic biases within the delivery mechanisms, including business models, market competition, regulatory framework or other factors affecting the operational delivery of services.

9.1.6. Insufficient variation in operational models

Many ICT services are owned and operated by a limited number of providers, often to near-monopoly levels. This limits choice of service provision and consequently user choice of the business model under which a service is provided. In many cases the business model is identical across all providers within an ICT sector, such that change of provider does not change the circumstances under which the user accesses the service. In many instances autonomy-preserving systems do not exist to the degree that a real alternative is available. In addition, intellectual property regulations and exploitative business models promote walled gardens and suppresses interoperability, further limiting user choice it is not possible to interoperate between providers.

9.1.7. Epistemic control

Epistemic control is the strategy of preventing user control of data flows at the interface. This is most commonly done to restrict the delivery of data or designing its delivery so as to hide information. A common example is the creation of privacy policies which are intentionally difficult to understand. Epistemic control can also be exercised by simply hiding what is being done, as we saw with elements of commercial user surveillance. Epistemic control can also be accomplished through interface design, whereby the presentation of data at the interface can influence the affordances through which it is understood by the user. More complex modes of epistemic control will become possible as systems become more sophisticated. For example, cultural bias in voice control systems can force people into modes of speech which are unwelcome to them. Epistemic control can also be achieved by "force feeding" data into the interface against the user's wishes.

9.1.8. Control of data processing

User autonomy can be restricted by constraints in data processing. Examples include blocking interoperability with a device from a competing manufacturer, controlling what apps can be installed on a device, or restricting devices from accessing certain files or formats. People can also be nudged into certain modes of data processing through design. For example, a device provider may embed systems into the device in a manner which makes it hard to avoid using them. Architectural and business models, most importantly cloud services and rentier capitalism (such as software-as-a-service) represent an attempt to own portions of the user's data processing and thus restrict their autonomy.

9.1.9. Control of standards

Control of standards is control of inter-operability. As we have seen, inter-operability frequently a necessary precondition of user choice. Restriction of user choice is one of the most common vectors for restrictions in autonomy. On this basis, commercial organisations desirous of, or in possession of, monopolies for provision of a service, must be regarded as antithetical to human autonomy, harmful to the development of best standards and should not be permitted a voice in the development of standards within computer science. The development of standards must be regarded as a fundamental foundation of future societies and subject to the same scrutiny and modes of development as are more traditional regulatory regimes, such as criminal law.

9.1.10. Summary – the problem is people, not systems

There is no necessity why any ICT should restrict human autonomy. Autonomy-preserving alternatives are always possible. Yet the current trends in ICT innovation are towards restrictions in autonomy, not liberation. This is the result of choices made by people, not the unavoidable by-product of functionality. These choices stem from ignorance, lack of thought, a desire to exploit people, a lack of respect for fundamental human rights and exploitative business models. Avoiding unpleasant futures requires changing how members of the computer science discipline think, disempowering restrictive regulatory regimes and promoting alternatives so that people have genuine choice in ICT services.

9.2. Answering the research questions

This section is a formal statement regarding how each research question was answered.

The primary research question was

"What are the possible threats to human autonomy generated by emerging ICT's?"

This was answered in *Threats to Autonomy from Emerging ICTs* (Dainow 2017b) (see *Chapter 6. Autonomy under Emerging ICT's, p.112*) and *Binding the Smart City Human-Digital System with Communicative Processes* (Dainow in-press) (see *Chapter 8. Autonomy within the Integrated Node, p. 148*).

Secondary questions were:

1. "What are the emerging ICT's?"

This was answered in *Threats to Autonomy from Emerging ICTs* (Dainow 2017b) (see *Chapter 6. Autonomy under Emerging ICT's, p.112*).

2. "What is human autonomy?"

This was answered in *Threats to Autonomy from Emerging ICTs* (Dainow 2017b) (see *Chapter 6. Autonomy under Emerging ICT's, p.112*).

3. "By what mechanism can an aspect of an emerging ICT have a restrictive effect on human autonomy?"

This was answered in *Key Dialectics in Cloud Services* (Dainow 2015b) (see *Chapter 4. The Current State of Affairs – Poles of the Debate, p.73*) and *Digital Alienation as the Foundation of Online Privacy Concerns* (Dainow 2015a) (see *Chapter 5. Digital Alienation - ICT's as Socio-Technical Systems, p. 91*)

9.3. Project Summary

This research project commenced by accepting the premise that technology is best defined as a socio-technical system. This definition combines three terms; 'social', 'technical' and 'system.' Each of these needs to be fully represented in any account of the interactions between ICT's and humans. We understand 'social' to require an account of the manner in which social factors influence, and are influenced by, the material and operational conditions of engineered artefacts and their functions. We understand the term 'technical' as referring to these engineered artefacts. Finally, we understand 'system' as requiring reference to principles of General Systems Theory as a meta-theory, and as requiring some reference to relevant domain-specific systems theories. In this case, the domain-specific theories must be relevant to the interactions between the social and the technical. The combination of these three terms means we must understand this relationship between the social and the technical as a system, not as a chain of linear causality. An account of each term was a necessary preliminary step before examining how emerging ICT's may impact human autonomy. In addition, it was necessary to define autonomy itself before it was possible to examine how it can be affected by emergent ICT's.

The following sections summarise how each issue was addressed.

9.3.1. ICT's as technical

Our research commenced with ETICA's findings, which also accepted the definition of technology as a socio-technical system. Here we encountered a detailed and reliable taxonomy of emerging ICT's defined by their technical characteristics. Our analysis of the ETICA project has not been published, nor has the methodology we developed by which to undertake this analysis. However, it is contained within this thesis (see *3. Evaluation of ETICA's Technology Analysis, p.33*). Our chief finding was that ETICA produced a sufficiently robust and detailed account of the technical aspects of emerging ICT's for use within our research project.

9.3.2. ICT's as social

Our analysis of the ETICA project found the work focused almost exclusively on the functional aspects of technology, with little analysis of social causation and inconsistent analysis of social impact. Furthermore, ETICA lacked a coherent, explicit or detailed social account, and no reference to ICT's as systems. While this did not undermine ETICA's own objectives, it rendered ETICA's technology identification insufficient for our research goals. We therefore developed an account of the social within the socio-technical. The aim here was not to develop a social account of technology in general, but merely one which could support an account of the impact on human autonomy in the future. We termed the generic state of having one's autonomy restricted by ICT's 'digital alienation.' Our account of the mechanisms by which digital alienation occurred was constructed through an analysis of marxist efforts to understand how current online systems generate ethically negative consequences for their users. Here we found marxist analysis mired in traditional structural elements of social analysis which rendered attempts to account for current circumstance unsatisfactory. Our position was that this failure derived from maintaining traditional terminology and theoretical forms in the face of a radically different environment from that under which Marx developed his theories. In the face of changed empirical circumstance, the only way to maintain terminology and the form of any theory is to redefine the terms. The central problem confronting marxist analysis of the internet is the rise of mechanisms by which value can be extracted from voluntary activity instead of paid labour. As a consequence, maintenance of form and terminology requires redefinition of 'labour' to include voluntary activity. With this redefinition, 'labour' comes to refer to almost any human activity and we lose the ability to distinguish between activities for pay and other human activities. Our approach was instead to maintain terminology but update the theoretical form to reflect the changed circumstances. We demonstrated that value is extracted from all forms of internet activity, so all that was required for value extraction was that one "exist" online. Replacing the term 'labour' with 'existence' was sufficient to account for modern circumstance while also maintaining the theoretical form. Our change therefore maintained coherence with the rest of the mechanisms within the marxist account and preserved its explanatory status.

Published as *Digital Alienation* (see *5. Digital Alienation - ICT's as Socio-Technical Systems, p. 91*), this account paired with its sister paper, *Key Dialectics* (see *4. The Current State of Affairs – Poles of the Debate, p. 73*), to add a critique of those dominant social and technical factors present in current circumstance which threaten human autonomy. Social factors covered included privacy violations, privacy regulation and business models. Technical factors included data structures (such as the digital shadow), inter-operability (or lack thereof), conformance to standards, algorithmic justice and interface design. In our view, an important aspect of these two papers is that both marxist and capitalist analyses reach the same conclusions regarding ethical impact and both agree on the causes. *Key Dialectics* also demonstrated a number of premises critical to our future research - that ICT's could create ethical problems, that personal data and personalisation could be vectors for ethical violations and that such circumstances were avoidable.

9.3.3. Defining the term 'autonomy'

While autonomy is understood as an enabling condition for most human rights, this understanding is not nuanced by the different accounts of what constitutes autonomy within modern philosophy. The general position outside philosophy is that humans have the capacity for autonomy, without reference to what that means beyond "self-governance." We took the position that any of the major accounts of autonomy could be legitimate if an individual chose to self-govern in accord with that account. We did not take a position that any one account was better or "more accurate" than any other. Consequently, we did not feel it appropriate to settle on any one definition of autonomy and restrict our analysis to it. Our position was that ICT's which prevented enaction of any version of autonomy restricted the autonomy of people who wished to live by that version.

By cross-referencing ETICA's technology groups against the different versions of autonomy, we demonstrated whether an ICT threatens autonomy usually depends on the version of autonomy someone is living by (see *6. Autonomy under Emerging ICT's, p.112*). We believe this is an important finding. It means that it can be the case that the same technological function can be both liberating and restrictive. It means that the impact on autonomy cannot be determined as a factor of the technology, but only by reference to specific individual modes of self-governance. In keeping with technology as socio-technical, technology's impact on autonomy can be understood only by considering the interaction of the social with the technical. This finding indicates the potential for future legal or ethical objections to technologies or operations which many people find unremarkable. For example, under versions of autonomy which emphasis relationships over self-sufficiency, it is possible to argue that the delivery of services by robots in medicine or education violates human rights. Combined with our analysis of the processes of digital alienation, what is revealed is that restrictions of human autonomy need not be the result of intentional efforts, but can derive from treating the user as a means to an end, from reification of only one form of autonomy in design, through business practices which restrict user choice or through pervasive behavioural nudging. The threat to human autonomy is thus insidious in that it is pervasive, hard to detect, and often requires complex explanation, understanding of which is dependent on specialist knowledge. Preventing it cannot be achieved through simplistic regulations or by reference to specific contexts or practices, but only through nuanced responses.

9.3.4. ICT's as systems

However, these conclusions do not explicitly address technology as a system. When considered as a system, ICT's cannot be divided into discrete groups because relevant processes span multiple technologies. Moreover, the impact on society derives more from the combined effects of ICT's than from any single one. In addition, the relationship between the social and the technical can only be understood in terms of systemic properties such as feedback and equifinality. It is thus possible to address the issue of the impact of emerging ICT's on autonomy from two perspectives. One is to address the impact each technology has on each version of autonomy. The other is to identify processes by which all IC technologies, viewed as a single socio-technical system, have the potential to restrict human autonomy. This second approach requires an account of a system which contains both social and technical elements. Processes by which autonomy is restricted must then be developed within the terms and structures of this system. Our account of the environment encapsulating the combination of all ICT's was drawn from existing work on the smart city. We applied two systems theories from sociology to develop a theoretical model of a truly socio-technical system, in which technical artefacts and humans hold equal causative status for each other and for the dynamics of the system (see *7. Understanding the Smart City as a System, p.132*). We were then able to describe mechanisms of autonomy restriction in terms of flows of energy and data through the system's nodes (see *8. Autonomy within the Integrated Node, p.148*).

9.3.5. Results - two complementary answers to the primary research question

We believe these two approaches to the research question are complementary. Our first answer[15] provides relatively detailed understanding of the manner in which specific ICT's impact specific versions of autonomy. Our second answer[16] identifies mechanisms of concern which can apply to any ICT or combination thereof, without variable import under different accounts of autonomy. We hold the latter account to be more in accord with the central interpretive principle of foresight studies – that the future cannot be predicted in any detail. It is therefore important that any foresight studies account focus on principles and processes which are not dependent on specific details about the future. While we find the ETICA findings compelling and valuable, they essentially depend on extrapolation of current trends into the future. This does not allow for unexpected or disruptive innovations, or for unanticipated changes in praxis.

[15] 'Threats to Autonomy from Emerging ICTs'. *Australasian Journal of Information Systems* 21, no. 0 (2017). https://doi.org/10.3127/ajis.v21i0.1438.

[16] 'Binding the Smart City Human-Digital System with Communicative Processes'. In *Technology and the City*, edited by Michael Nagenborg, in-press.

The first analysis is valuable for providing a detailed account of the technologies currently emergent. However, the second account can be applied to other technologies which have yet to be invented. More importantly, it provides axioms by which potential systems can be evaluated during their initial design phase. We believe these two accounts provide a comprehensive understanding of the manner in which emerging ICT's can threaten human autonomy in the future.

9.4. Summary of Contributions to Knowledge

This research offers a number of contributions. These were detailed above (see *1.5 Contributions, p.14*), and are summarised here:

9.4.1. Resolving difficulties in marxist analysis of the digital economy

The digital economy does not generate value in the same way as manufacturing did when Marx developed his theories because of its extraction of value from voluntary and unpaid activity. As a result, traditional marxism cannot account for digital alienation. By the simple step of replacing labour with existence, together with empirical evidence supporting this move, we resolve marxism's inability to account for the digital economy.

9.4.2. Development of a formal, generic, methodology for evaluation of large foresight research projects which combine multiple methodologies.

Significant foresight research projects commonly combine multiple methodologies and can arise in many disciplines. Evaluation of such projects is therefore problematic. This is a widely recognised issue in foresight studies. Our adaptation of medicine's CASP provides a formal and detailed methodology for analysis of such foresight research. We do not believe a methodology of equivalent rigor or depth exists within the field. We therefore believe that this evaluation method can have widespread value.

9.4.3. A catalogue of potential threats to human autonomy across the range of emergent technologies.

Threats to Autonomy from Emerging ICT's (Dainow 2017b) is the first time any analysis of autonomy under ICT's has considered the different versions of autonomy. Any ethical analysis of ICT's which is concerned with human rights or other values which derive from autonomy must consider the different versions of autonomy or stand a good chance of neglecting those who live according to values different from the version of autonomy used in the analysis. This is particularly important when drafting legislation and in any other efforts to maintain human autonomy into the future.

9.4.4. Development of Integrated Domain Theory

Integrated Domain Theory introduces important concepts to ethical analysis of digitally-driven environments. It provides a means of discussing complex situations, composed of multiple systems and people, in depth, but without becoming overwhelmed by detail. It provides a means of identifying common ethical issues which can occur in multiple contexts, especially those deriving from system architecture. The approach which gave rise to Integrated Domain Theory is General Systems Theory and derivatives. Applications of systems theory have already demonstrated value in other fields, especially sociology, biology and geography. Our contribution has been bring it into ICT ethics by adding features unique to the discipline.

10. Limitations, Recommendations and Future Work

10.1. Limitations

10.1.1. Arguments against a "correct" version of autonomy

Philosophers who hold there is only one legitimate version of autonomy and all others are wrong will argue that failing to select one account of what constitutes human autonomy is a serious limitation to the study. From that perspective, seeking to analyse how technologies affect different versions of autonomy would not be an appropriate approach. The most significant limitation to this work is that we have not published a formal argument in favour of the legitimacy of multiple versions of autonomy. The lack of such a formal argument leaves us defenceless against arguments that a particular form of autonomy is the only correct one. However, while many have proposed their own accounts of what constitutes autonomy, often rejecting competing versions, our research has not revealed any formal arguments supporting the position that there must be a single correct account of what constitutes autonomy. Instead, this seems to be an assumption.

However, this does not undermine the analyses we have published. *Threats to Autonomy from Emerging ICT's* cross-references technologies with all major accounts of autonomy. Whether one agrees or not with a particular definition does not affect how that definition interacts with a particular technology. *Binding the Smart City* intentionally considers autonomy in a generic fashion, such that its analysis is not changed by using any particular account of what autonomy is.

10.1.2. English language

One limitation to this research is that it is drawn purely from English-language sources. Some of the most important sources used, such as Latour, Kant and Luhmann, were written in French or German, but in all cases we have drawn on English translations. We have been sensitive to issues of translation, particularly the danger that critical terms may not have an exact equivalent in English. This has not been a frequent issue, but where it has occurred we have attempted to cross-reference multiple translations as well as appropriate dictionaries. We do not think this has compromised our own work, though specialists may disagree with our understanding of minor nuances of some theories.

10.1.3. Western thinking

We are aware that our perspective and approach is particularly Western and bound in certain ways of thinking which are not reflected in other philosophical cultures. For example, Daoist philosophical approaches in China and the native Quechuan philosophies of Latin America do not treat dialectical opposites as being in conflict, as they are treated in the West. This affects our thinking about binary choices and directly opposing values in that it promotes an approach of seeking to choose between them or fuse them into a new synthesis. In contrast, the Daoist or Quechuan approach would promote viewing the tension between them as the important dynamic to be analysed and used, not something to be resolved. While we recognise that our work is bounded by the limits of Western and Enlightenment approaches, we do not think this has compromised the analysis. We have remained consistent within the framework of Western philosophical analysis and so our conclusions remain coherent with the premises and the mode of analysis. It is possible approaches from other philosophical cultures could expand our understanding, but the internal coherence of our analysis means including other perspectives would not undermine it.

10.1.4. Prioritising autonomy

We are aware that in thinking about the preservation of autonomy, we did not contrast it with other values. Given that autonomy is the subject of our research we must prioritise it within our analysis, but this may not reflect its relative priority against other values, such as communitarianism and paternalism. Furthermore, no attempt was made to contrast restrictions on autonomy with compensating ethical benefits, such as increased social harmony. Many cultures give greater priority to the group over the individual than is common in the West, which means different cultural approaches will regard the import of our findings differently. However, this does not undermine our analysis of how autonomy can be restricted. It merely means we have not sought to counter-balance any restriction of autonomy with any potential benefit. This was an intentional element of the research design. The research question was intentionally focused purely on how autonomy could be restricted, not the import or context of such restrictions, which would have considerably complicated the research process. In addition, compensatory benefits in the future are even more speculative and difficult to predict. Attempting to do so would have gone beyond what can be reasonably undertaken under the principles of future studies. Furthermore, the degree to which any benefit compensates for loss of autonomy depends on the version of autonomy in question and the relative value of other ethical virtues. We believe our research has wider accessibility because it does not pre-suppose either and is therefore accessible to proponents of any version of autonomy and any compensatory values.

10.2. Recommendations

10.2.1. Regulations should remain vague about autonomy

The most important recommendation derives from competing versions of autonomy, all of which we hold to be legitimate. We cannot therefore base rules and regulations, especially those regarding consent, on any specific version of autonomy. To do so is necessarily to reduce the autonomy of those who choose to live by different rules. Though we have not formal arguments in support of our position, we hold that to argue one version of autonomy is legitimate and all others are not is inappropriate because autonomy is a created concept and there is therefore no empirical test of that to which the term refers[17]. Furthermore, we cannot refer back to the original conception developed by Kant because he is strictly limited his definition to specific procedures in moral reasoning. This offers little for our understanding of political or personal autonomy. Laws cannot therefore be based on a particular version of autonomy. This does not seem to be a significant concern so far because, while laws and regulations make reference to autonomy, they generally do so in such a diffuse fashion which allows for most accounts of autonomy. It is essential that this pattern continue and that regulation regarding ICT's not pin itself to specific aspects of a single conception of autonomy.

10.2.2. Computer science needs to develop ethical awareness

Once a science starts to produce effects of ethical significance in people's lives it must develop an ethical aspect, both in terms of practical methodology and in terms of moral virtues. The mind-set of the computer scientist involved in the development of ethically significant technologies has been, and continues to be, ethically ignorant. This will change. Society will not tolerate the development of socially harmful technologies forever. Sooner or later it will enforce an ethical awareness upon computer science, just

[17] Our position is that the appropriate issue for philosophy is not "what is the nature of autonomy?" but rather the following questions - "what justifies an individual's claim to be acting autonomously and how must they be treated as a consequence by others?" and "by what warrant may an individual state that another is not acting autonomously when they claim they are so doing?"

as it did with Biology and Medicine. It would be preferable, more productive and better for all, if the field of computer science voluntarily embraced ethical education and ethical awareness as virtues to the same degree that the field embraces other scientific virtues, such as intellectual honesty and methodological consistency. This requires the inclusion of ethical education throughout computer science education. It further requires encouragement of ethical awareness and ethical aspiration within the graduate profession, both in universities and in industry. In that a significant number of ICT innovations are developed by people who did not complete their first degree, it is probably best that ethical education commence in computer science on the first day of class. At this time we do not have canonical methodologies for teaching ethics within computer science. A very limited number of textbooks have been attempted and a very limited amount of research has been conducted. These constitute starting points, but we cannot expect to roll out a fully developed educational component for ethics within computer science immediately. Research needs to be conducted to determine the best methods by which to develop ethical awareness in computer science. Such research should commence with an examination of how ethics has been introduced into other sciences, most notably medicine, and practical experimentation in the classroom.

10.2.3. Computer science education should include the humanities

ICT systems which interact with human beings will suit people better, and are less likely to be ethically problematic, if they are based upon a more sophisticated understanding of human beings and society. If computer scientists are going to build systems which become part of people's personal world, they need to understand those people and their personal world in order to achieve the best fit between what they create and what people need. We cannot assume, as we do now, that this will happen magically through some form of osmosis by students during an education which contains no such subject matter. At best we are leaving it to chance, relying on individuals to do their own reading and personal development. The repeated ethical debates which arise regarding ICT systems indicate students will not independently develop such an awareness to the degree that it will affect the paths of future innovation. It would therefore seem appropriate to demand computer science students take some training in the humanities. Computer Scientists need to have an understanding of the culture within which they are building the systems as well as the people. This will become increasingly important as computer science systems spread into civic management and planning systems and so begins to have a greater effect on social processes.

10.2.4. Maximise user-level configuration of digital systems

In order to permit people to live in an intimate and personalised relationship with their digital environment, while also being able to exercise a variety of ways of living autonomously, the systems in which people are embedded must be configurable to that individual's style of autonomy. This demands that a loose conception of autonomy be embedded in digital systems by the designers so as to avoid imposing restrictions on user autonomy. Instead, systems should permit the user to configure the systems they use with the widest possible variation of operational characteristics, so that the individual's digital environment supports a style of life chosen by the user. Such user configuration must include all the elements we have identified in this research as capable of threatening autonomy. Unless personalised digital environments are configurable to the user's satisfaction, such that they permit the user to exercise autonomy as they see fit, those systems will necessarily limit the autonomy of some, if not all.

This demands that digital systems offer the following:

1. Variations in business model and terms of service.
2. Inter-operability between competing services, such that a user may move between providers without loss of functionality.
3. Limitations to, and user control of, the degree and content of digital surveillance and the use thereof.
4. Where viable, user-selectable alternatives in system architecture. For example, no one should be forced to use cloud-based services if local processing is a viable alternative. In fact, we would advocate the legal mandating of use of Langheinrich's Principles for Privacy-Aware Ubiquitous Systems throughout ICT innovation, only using cloud-based services when no other architecture is possible.
5. Increased restrictions in current intellectual property law and other regulations which support software-as-a-service and rentier capitalism.

10.2.5. Special legal and ethical status for personal digital devices

The chapter on the integrated node made brief reference to integrated devices as having special status because of their intimate connection with their user's personal lifeworld (see *8.1.3 The Integrated Node, p.157*). We need a new legal status for such personal digital devices because their relationship to the individual is significantly different from other property. The same special status should also be accorded to any other devices through which we interact with our digital environment on a personal level, or which are essential for maintenance of normal life. Because of their essential nature they are necessary for the maintenance of the user's human dignity. People therefore have a human right to own and operate these devices in a manner which does not undermine their autonomy. The degree to which the device is personalised to the user is proportional to the degree to which these devices are deserving of a special ethical and legal status. In particular, devices which are embedded within the body or essential for continued existence must be the unique and special property of the person within whom they reside in any and all respects. Restricting ownership or use of such devices has the same impact as limiting a person's use of their own body. We therefore recommend commencing determination of the details of such special status by viewing such personal devices as analogous to the human body, rather than to human possessions, as has been the case until now.

10.3. Further work

10.3.1. Maintenance of ETICA

The technological taxonomy developed by the ETICA project remains a unique and valuable resource. It should be more widely disseminated through a dedicated public website. The taxonomy should be updated as soon as feasible and a long-term project put in place to keep it updated on a regular basis, probably every five years. It is not necessary to recover ETICA's lost data or use identical methodologies as were used before. In particular, we would not recommend the use of the same tools for bibliometric analysis. More sophisticated tools exist today and the methodologies need to adapt as new technologies develop. An ETICA website and taxonomy should form the spine of a repository of discussions around these technologies.

10.3.2. Enhancing our understanding of autonomy

Much further work needs to be done with regard to the various versions of autonomy and their interaction with ICT's, especially with regard to personalisation and interface design. In particular, research needs to be done to provide a much more detailed taxonomy of autonomy, perhaps similar in format and approach to ETICA's analysis of technologies. Furthermore, it is likely that similar examination of other fundamental ethical values such as freedom will reveal multiple, competing versions, each of which need to be cross-referenced with ETICA's technology groups.

10.3.3. Development of Integrated Domain Theory

We have found Integrated Domain Theory to be a productive application of systems theory to smart city analysis and the analysis of any complex digital environment. Much further work remains available, such as the development of basic metrics for the calculation of energy flow between nodes. Concepts suited to the development of formal metrics include thresholds and resistance, transduction between informational and physical input and output, configurability and variations in flow patterns. For example, that digital environments will accept as input and produce as output both informational and physical energy suggests the possibility for the development of metrics and calculations for the interaction between physical and informational energetic structures.

Under Integrated Domain Theory, digital systems are considered equally important causative agents to humans. With the increase in artificial intelligence and autonomous decision-making, we see the development of a "digital perspective." Integrated Domain Theory provides a means of exploring the nature of this perspective and its impact on human life. Much of the initial work can be done through logical deduction of axioms. For example, in order for a system to deliver a personalised service, it must have access to information about the person. There must, therefore, be a body of data about each person who receives personalised services. The accuracy and granularity of that data will significantly influence the degree to which service delivery suits user need. It is our intention to continue development of Integrated Domain Theory and we have already identified twelve axioms similar to this example which define the digital perspective.

11. Final Reflections

This section contain informal reflections in which I seek to contextualise the research within the bigger frames of human history and theology.

My understanding of this research project positions it in a profound change occurring today, sometimes known as the "fourth industrial revolution" (Schwab 2016). We are moving into a stage of history where our external environment will be intelligent and autonomous, determining for itself how it responds to us. There is nothing inevitable about this direction or its endpoint; it will be the product of human choices. As a society we are not making those choices. They are being made for us by corporations in pursuit of their own aims. There is no reason to assume they will create a digital environment for us which is tuned to our needs to the best possible degree. Current trends give us every reason to expect this environment will be tuned to using us as a resource. This is not inevitable and can be changed with relative ease through appropriate carrots and sticks. The most important is the prevention of monopolies; whether that monopoly be of a service, a particular sphere of activity or a geographic space. There is no technical reason why we cannot mandate a variety of technical architectures and business models for every place and purpose in our digital environment. Whether considered from a marxist or capitalist perspective, this would be a better environment for society and the individual.

The emerging digital environment is coming to interpenetrate the human body. Conversely, the human self, through cognitive offloading, prosthetics and embodied devices, is spreading into the built environment. Therefore we cannot formulate our policies or ethics through a model based on property as being physical objects which exist outside the body. We are thus entering a new of being characterised by the inseparable fusion of the digital and the human. This new way of being does not fit into pre-existing conceptions of what is inside and what is outside. There now exist external objects which we treat as part of ourselves, which the individual sees as cognitive extensions of themselves, and there are elements of the external world within our bodies, deeply embedded within our private personal lives. This new zone of overlap between the person and their environment requires recognition, definition and analysis. It requires a new way of understanding our relationship with our technology - in which we do not dominate tools, but neither do they dominate us. Instead both are subparts of something which can only be understood as a fused being - human and digital – which I have termed the Integrated Personage. The human is not dominant in this being, but rather is dominated by something new, the unification of the two. This Integrated Personage, being a fusion of two different perspectives, is incomprehensible to the human. We may imagine, describe and organise our vision of the digital perspective, but we cannot know it. Accordingly, that aspect of the experience of the Integrated Personage is beyond human comprehension. It is greater than us. We are each becoming part of something bigger than ourselves.

From a historical point of view we are moving into an age where it is possible to have control of the individual and the environment to a degree never seen before in human history. This represents a quantum leap in the power to control people. We do not have the legal or organisational checks in place to deal with such a unprecedented situation. Furthermore, most leaders are ignorant of the potential dangers. Indeed, the potential power is so great it challenges anyone's imagination. Yet none of the dangers are unpredictable and, in most cases, they are merely magnified examples of situations we have seen in the past history of technology. We know how to deal with these dangers. In this age all that is required is that we look and act. Autonomy cannot be restricted by ICT's unless those ICT's, as sociotechnical systems, exhibit certain properties. These properties have been detailed throughout this thesis. They include walled silos, monopoly service provision, privacy violations, insufficient user configurability, poor design and absurdly naive understanding of human beings and society. What is required is education of users and developers, appropriate regulatory mechanisms to prevent certain structural features emerging in the marketplace and service provision, and the development of theological, ethical and legal responses to the rise of the Integrated Domain, integrated devices and the Integrated Personage.

I believe it is inevitable that this will happen, as it always has in the past. However, the historical pattern is usually that we have to have disasters before we will recognise the danger and respond. In the broad sweep of history those disasters have not been limited to relatively short timescales, but have also been constituted of significant declines in the quality of life and the freedom of the individual for hundreds of years. There is nothing inevitable about the continuing rise of human liberty and fulfilment, nothing in the nature of society or the human which can prevent declines that last for ages. Amongst the other dangers facing the human race at this time there now exists the possibility that monopoly service providers could, through extension of current legal regimes and directions in innovation, turn us all into digital slaves, partially or fully owned by them, with little or no freedom.

A particular concern is the Chinese social credit system. If this system fulfils its aims it will serve as a template for any dictatorship which wishes total control of the population. The Chinese social credit system may be so successful as to prevent resistance and the rise of freedom in China for hundreds of years. If that is the case, over the course of time, it will inevitably spread. As each nation goes through those dark times which are almost inevitable in its history, and is taken over by dictatorship, the use of a successful totalitarian control system like the Chinese social credit system could preserve dictatorial governments until one day most, if not all, human beings live under such control. Is such a state likely? We simply do not know.

Most readers will have an immediate emotional reaction that this is inconceivable, that something will always happen to prevent it. However, that is merely an emotional reaction and is not based on any empirical knowledge. We are therefore not in a position to gauge the danger we are in as a society. What we are in a position to do is undertake steps which we know will not only reduce the risk, but also create better systems - which work better and serve people and society more effectively, more efficiently and more healthily. Not to see this as preferable, as an aim appropriate to computer science, as a desire to produce the best systems with the closest fit to the design goal, not to see this as the ultimate objective of the application of computer science is to be a poor computer scientist. It must be reasonable for any well-adjusted human being to seek that their work, at least to do no harm, if not make the world a better place.

There is nothing incompatible about an ethical concern for the products of one's science and the rigorous pursuit of that science to the best of one's ability. To position ethics as external to computer science is either a failure to understand what is meant by scientific objectivity or simply amoral. Scientific objectivity does not call for a lack of ethics. Scientific objectivity calls for truthful adherence to results. The framework within which research is done is ethically significant and does not inhibit good science. It is time for computer science to understand that the applications of the discipline are now central to everyday human life and the future of society. Computer science cannot be regarded as a purely theoretical or abstract field disconnected from ethics. Even the most theoretical and abstract research today will result in real-world products which affect people's lives at some stage in the future. It is time for computer scientists to decide whether they wish to ignore this issue until society forces regulation upon them or whether they wish to develop their own ethical awareness before being forced.

I hope that my research has given computer science tools and concepts with which to adapt the direction of ICT innovation (and the supporting science behind it) in a positive fashion and avoid the disasters which my research shows could occur.

12. Bibliography

Aaltonen, Milka, and Theodor Barth. 2005. 'How Do We Make Sense of the Future?' *Journal of Future Studies* 9 (May): 45–60.

Abse, Nathan. 2012. 'Microtargeted Political Advertising in Election 2012'. IAB Presents Innovations in Web Marketing and Advertising. New York, NY, USA: Internet Advertising Bureau. http://www.iab.net/media/file/Innovations_In_Web_Marketing_and_Advertising_delivery.pdf.

Acar, Güneş, Brendan Van Alsenoy, Frank Piessens, Claudia Diaz, and Bart Preneel. 2015. 'Facebook Tracking through Social Plugins'. Brussels: Belgian Privacy Commission.

Ackoff, Russell L. 1989. 'From Data to Wisdom'. *Journal of Applied Systems Analysis* 16 (1): 3–9.

Acquisti, Alessandro. 2009. 'Nudging Privacy: The Behavioral Economics of Personal Information'. *IEEE Security & Privacy* 7 (6).

Acquisti, Alessandro, Allan Friedman, and Rahul Telang. 2006. 'Is There a Cost to Privacy Breaches? An Event Study'. In *Proceedings of the Twenty-Seventh International Conference on Information Systems (Citeseerx.Ist.Psu.Edu/Viewdoc/Download?Doi=10.1.1.73.2942&rep=rep1&type=pdf.*

Addington, D Michelle, and Daniel L Schodek. 2005. *Smart Materials and New Technologies*. Oxford; New York: Routledge.

Afroz, Sadia, Aylin Caliskan Islam, Jordan Santell, Aaron Chapin, and Rachel Greenstadt. 2013. 'How Privacy Flaws Affect Consumer Perception'. In , 10–17. IEEE. https://doi.org/10.1109/STAST.2013.13.

Agich, George J. 2003. *Dependence and Autonomy in Old Age: An Ethical Framework for Long-Term Care*. 2nd ed., rev. Cambridge, UK ; New York: Cambridge University Press.

Ahuvia, Aaron C. 2005. 'Beyond the Extended Self: Loved Objects and Consumers' Identity Narratives'. *Journal of Consumer Research* 32 (1): 171–184.

Airaksinen, Timo. 1988. 'An Analysis of Coercion'. *Journal of Peace Research* 25 (3): 213–27.

Albino, Vito, Umberto Berardi, and Rosa Maria Dangelico. 2015. 'Smart Cities: Definitions, Dimensions, Performance, and Initiatives'. *Journal of Urban Technology* 22 (1): 3–21. https://doi.org/10.1080/10630732.2014.942092.

Albrechtslund, Anders. 2007. 'Ethics and Technology Design'. *Ethics and Information Technology* 9 (1): 63–72. https://doi.org/10.1007/s10676-006-9129-8.

Amazon. 2017. 'Kindle Terms of Use'. 2017. http://www.amazon.com/gp/help/customer/display.html?nodeId=200506200.

Amichai-Hamburger, Yair, Galit Wainapel, and Shaul Fox. 2002. '"On the Internet No One Knows I'm an Introvert": Extroversion, Neuroticism, and Internet Interaction'. *CyberPsychology & Behavior* 5 (2): 125–28. https://doi.org/10.1089/109493102753770507.

Anderson, D.P., and G. Fedak. 2006. 'The Computational and Storage Potential of Volunteer Computing'. In *CCGRID '06 Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid*, 73–80. Washington, D.C.: IEEE Computer Society. https://doi.org/10.1109/CCGRID.2006.101.

Anderson, Joel. 2014. 'Regimes of Autonomy'. *Ethical Theory and Moral Practice* 17 (3): 355–68. https://doi.org/10.1007/s10677-013-9448-x.

Anderson, Joel, and John Christman. 2005. 'Introduction'. In *Autonomy and the Challenges of Liberalism: New Essays*, 1–23. Cambridge, UK; New York: Cambridge University Press.

Anderson, Scott. 2008. 'Of Theories of Coercion, Two Axes, and the Importance of the Coercer'. *Journal of Moral Philosophy* 5 (3): 394–422. https://doi.org/10.1163/174552408X369736.

———. 2010. 'The Enforcement Approach to Coercion'. *Journal of Ethics and Social Philosophy* 5 (1).

Andrejevic, Mark B. 2011a. 'Estrangement 2.0'. *World Picture* 6: 1–14.

———. 2011b. 'Surveillance and Alienation in the Online Economy'. *Surveillance & Society* 8 (3): 278–87.

———. 2012. 'Exploitation in the Data Mine'. In *Internet and Surveillance*, edited by Christian Fuchs, Marisol Sandoval, Boersma Kees, and Anders Albrechtslund, 71–88. Routledge Studies in Science, Technology and Society. New York, N.Y: Routledge.

Angwin, Julia. 2014. *Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance*. New York, N.Y: Henry Holt & Co.

Aquinas, Thomas. 1920. *Summa Theologica*. Translated by Fathers of the English Dominican Province. Second and Revised Edition, 1920. London: Fathers of the English Dominican Province.

Aristotle. 2014a. 'Metaphysics'. In *Complete Works of Aristotle, Digital Edition: Revised Oxford Translation*, edited by Jonathan Barnes, 1552–1728. Princeton, NJ: Princeton University Press.

———. 2014b. 'Posterior Analytics'. In *Complete Works of Aristotle, Digital Edition: Revised Oxford Translation*, edited by Jonathan Barnes, 114–66. Princeton, NJ: Princeton University Press.

Assink, Marnix. 2006. 'Inhibitors of Disruptive Innovation Capability: A Conceptual Model'. *European Journal of Innovation Management* 9 (2): 215–233.

Atzori, Luigi, Antonio Iera, Giacomo Morabito, and Michele Nitti. 2012. 'The Social Internet of Things (Siot)–When Social Networks Meet the Internet of Things: Concept, Architecture and Network Characterization'. *Computer Networks* 56 (16): 3594–3608.

Augusto, Juan Carlos, Hideyuki Nakashima, and Hamid Aghajan. 2010. 'Ambient Intelligence and Smart Environments: A State of the Art'. In *Handbook of Ambient Intelligence and Smart Environments*, edited by Juan Carlos Augusto, Hideyuki Nakashima, and Hamid Aghajan, 3–31. Springer.

Badley, Graham. 2009. 'Publish and Be Doctor-Rated: The PhD by Published Work'. *Quality Assurance in Education* 17 (4): 331–342.

Balakrishna, C. 2012. 'Enabling Technologies for Smart City Services and Applications'. In *2012 Sixth International Conference on Next Generation Mobile Applications, Services and Technologies*, 223–27. https://doi.org/10.1109/NGMAST.2012.51.

Balakrishnan, J.D., and J.A. MacDonald. 2001. 'Signal Detection Theory'. In *International Encyclopedia of Ergonomics and Human Factors*, edited by Waldemar Karwowski, 3:542–45. London; New York: Taylor & Francis.

Barrett, Lisa Feldman, Maria Gendron, and Yang-Ming Huang. 2009. 'Do Discrete Emotions Exist?' *Philosophical Psychology* 22 (4): 427–37. https://doi.org/10.1080/09515080903153634.

Barroso, Luiz André. 2005. 'The Price of Performance'. *Queue* 3 (7): 48–53. https://doi.org/10.1145/1095408.1095420.

Barry, Simon. 2012. *Continuum*. TV Series. Reunion Pictures.

Bass, Katy Glenn. 2013. 'Chilling Effects: NSA Surveillance Drives U.S. Writers to Self-Censor'. PEN American Center. http://www.pen.org/sites/default/files/Chilling%20Effects_PEN%20American.pdf.

Batty, M., K. W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali. 2012. 'Smart Cities of the Future'. *The European Physical Journal Special Topics* 214 (1): 481–518. https://doi.org/10.1140/epjst/e2012-01703-3.

Bedigian, Louis. 2016. 'TU Automotive Detroit 2016 Conference Report'. Detroit: TU-Automotive Ltd (Penton).

Belk, Russell W. 1988. 'Possessions and the Extended Self'. *Journal of Consumer Research* 15 (2): 139–168.

———. 2013. 'Extended Self in a Digital World'. *Journal of Consumer Research* 40 (3): 477–500.

Bell, D. 2014. 'Communitarianism'. In *The Stanford Encyclopaedia of Philosophy*, edited by Edward N. Zalta, Summer 2014. http://plato.stanford.edu/archives/sum2014/entries/communitarianism/.

Bell, Wendell. 1996. *Foundations of Future Studies*. New Jersey: Transaction Publishers.

Belleflamme, Paul, Thomas Lambert, and Armin Schwienbacher. 2014. 'Crowdfunding: Tapping the Right Crowd'. *Journal of Business Venturing* 29 (5): 585–609. https://doi.org/10.1016/j.jbusvent.2013.07.003.

Bergkamp, Lucas. 2002. 'The Privacy Fallacy'. *Computer Law & Security Report* 18 (1): 31–47.

Bernays, Edward L. 1928. *Propaganda*. New York, N.Y: Ig Publishing.

Berners-Lee, Tim. 1999. *Weaving the Web*. New York, N.Y: HarperCollins.

Bicocchi, Nicola, Alket Cecaj, Damiano Fontana, Marco Mamei, Andrea Sassi, and Franco Zambonelli. 2013. 'Collective Awareness for Human-Ict Collaboration in Smart Cities'. In *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2013 IEEE 22nd International Workshop On*, 3–8. IEEE.

Bicocchi, Nicola, Letizia Leonardi, and Franco Zambonelii. 2015. 'Software-Intensive Systems for Smart Cities: From Ensembles to Superorganisms'. In *Software, Services, and Systems: Essays Dedicated to Martin Wirsing*, edited by Rocco De Nicola and Rolf Hennicker, 538–551. London; New York: Springer. http://dx.doi.org/10.1007/978-3-319-15545-6_31.

Bittner, Rüdiger. 2014. 'Autonomy Modest'. *Erkenntnis* 79 (7): 1329–39.

Bogliacino, Francesco, and Mario Pianta. 2013. 'Profits, R&D, and Innovation—a Model and a Test'. *Industrial and Corporate Change* 22 (3): 649–78. https://doi.org/10.1093/icc/dts028.

Boldrini, Chiara, Marco Conti, Franca Delmastro, and Andrea Passarella. 2010. 'Context- and Social-Aware Middleware for Opportunistic Networks'. *Journal of Network and Computer Applications* 33 (5): 525–541.

Bonomi, Flavio, Rodolfo Milito, Jiang Zhu, and Sateesh Addepalli. 2012. 'Fog Computing and Its Role in the Internet of Things'. In *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*, 13–16. ACM.

Botta, Alessio, Walter de Donato, Valerio Persico, and Antonio Pescapé. 2016. 'Integration of Cloud Computing and Internet of Things: A Survey'. *Future Generation Computer Systems* 56: 684–700.

Boulding, Kenneth E. 1956. 'General Systems Theory-The Skeleton of Science'. *Management Science* 2 (3): 197–208.

Bourdieu, Pierre. 1986. 'The Forms of Capital'. In *Handbook of Theory and Research for the Sociology of Education*, edited by John G. Richarson, 46–58. Westport, CT.: Greenwood Publishing Group.

———. 1990. *The Logic of Practice*. Translated by Richard Nice. Stanford: Stanford University Press.

Bowerman, B, J Braverman, J Taylor, H Todosow, and U Von Wimmersperg. 2000. 'The Vision of a Smart City'. In *2nd International Life Extension Technology Workshop, Paris*. Vol. 28.

Brey, Philip. 2005. 'Artifacts as Social Agents'. In *Inside the Politics of Technology: Agency and Normativity in the Co-Production of Technology and Society*, edited by Hans Harbers, 61–84.

———. 2012. 'Anticipatory Ethics for Emerging Technologies'. *NanoEthics* 6 (1): 1–13. https://doi.org/10.1007/s11569-012-0141-7.

Brooks, Nathan. 2005. 'Data Brokers: Background and Industry Overview'. CRS Report for Congress RS22137. Washington, D.C.: Congressional Research Service, The Library of Congress.

Brown, Ian, and Christopher T. Marsden. 2013a. *Regulating Code: Good Governance and Better Regulation in the Information Age*. Information Revolution and Global Politics. Cambridge, Mass: The MIT Press.

———. 2013b. 'Regulating Code: Towards Prosumer Law?' *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2224263.

———. 2013c. 'Interoperability as a Standard-Based ICT Competition Remedy'. In *8th International Conference on Standardization and Innovation in Information Technology 2013*, 1–8. Sophia-Antipolis: IEEE. https://doi.org/10.1109/SIIT.2013.6774570.

Bublitz, Jan Christoph, and Reinhard Merkel. 2009. 'Autonomy and Authenticity of Enhanced Personality Traits'. *Bioethics* 23 (6): 360–374.

Burmeister, Oliver K. 2016. 'The Development of Assistive Dementia Technology That Accounts for the Values of Those Affected by Its Use'. *Ethics and Information Technology*, 1–14. https://doi.org/10.1007/s10676-016-9404-2.

Buss, Sarah, and Edward N. Zalta. 2015. 'Personal Autonomy'. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/personal-autonomy/.

Camp, L Jean, and Carlos A Osorio. 2003. 'Privacy-Enhancing Technologies for Internet Commerce'. *Trust in the Network Economy*, 317–331.

Cannataci, Joseph. 2016. 'Report of the UN Special Rapporteur on the Right to Privacy'. ADVANCE UNEDITED VERSION A/HRC/31/64. Human Rights Council.

Capurro, Rafael. 2009. 'Intercultural Information Ethics: Foundations and Applications'. *Signo y Pensamiento* 28 (55): 66–79.

Cardwell, Donald. 1994. *The Fontana History of Technology*. London; New York: Fontana.

Castoriadis, C., and K. Blamey. 1997. *The Imaginary Institution of Society*. MIT Press.

Chirkov, Valery I., Richard M. Ryan, and Kennon M. Sheldon. 2011. 'The Struggle for Happiness and Autonomy in Cultural and Personal Contexts: An Overview'. In *Human Autonomy in Cross-Cultural Context*, edited by Valery I. Chirkov, Richard M. Ryan, and Kennon M. Sheldon, 1:1–21. Cross-Cultural Advancements in Positive Psychology. Dordrecht: Springer Netherlands. http://link.springer.com/10.1007/978-90-481-9667-8.

Chourabi, Hafedh, Taewoo Nam, Shawn Walker, J Ramon Gil-Garcia, Sehl Mellouli, Karine Nahon, Theresa A Pardo, and Hans Jochen Scholl. 2012. 'Understanding Smart Cities: An Integrative Framework'. In *System Science (HICSS), 2012 45th Hawaii International Conference On*, 2289–2297. IEEE.

Christensen, Bill. 2009. 'ETICA Already Missed These SF Comm Predictions: Science Fiction in the News'. Technovelgy.Com. 16 April 2009. http://www.technovelgy.com/ct/Science-Fiction-News.asp?NewsNum=2242.

Christman, John. 2004. 'Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves'. *Philosophical Studies* 117 (1): 143–164.

———. 2014. ‘Relational Autonomy and the Social Dynamics of Paternalism’. *Ethical Theory and Moral Practice* 17 (3): 369–82. https://doi.org/10.1007/s10677-013-9449-9.

———. 2015. ‘Autonomy in Moral and Political Philosophy’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2015. http://plato.stanford.edu/archives/spr2015/entries/autonomy-moral/.

Cirillo, Marcello, Lars Karlsson, and Alessandro Saffiotti. 2012. ‘Human-Aware Planning for Robots Embedded in Ambient Ecologies’. *Pervasive and Mobile Computing* 8 (4): 542–561.

Cirucci, Angela M. 2015. ‘Facebook’s Affordances, Visible Culture, and Anti-Anonymity’. In *Proceedings of the 2015 International Conference on Social Media & Society*, 11:1–11:5. SMSociety ’15. New York, NY, USA: ACM. https://doi.org/10.1145/2789187.2789202.

Citron, Danielle Keats, and Frank Pasquale. 2014. ‘The Scored Society’. *Washington Law Review* 89 (1): 1–33.

Clarke, Steve. 2005. ‘Future Technologies, Dystopic Futures and the Precautionary Principle’. *Ethics and Information Technology* 7 (3): 121–126.

Clifton, Judith, Pierre Lanthier, and Harm Schröter. 2011. ‘Regulating and Deregulating the Public Utilities 1830–2010’. *Business History* 53 (5): 659–672.

Coates, Joseph F, and Jerome C Glenn. 2009. ‘Normative Forecasting’. In *Futures Research Methods*, edited by Theodore J Gordon and Jerome C Glenn, 3rd ed. Washington, DC: The Millenium Project.

Cobham Plc. 2011. ‘Cobham Tactical Communications & Surveillance’. Brochure. Cobham Plc Divisions & Capabilities. Dartmouth, Canada: Cobham PLC.

———. 2012. ‘Cobham Smart Cities Surveillance Architecture’. Brochure. Dartmouth, Canada: Cobham PLC.

‘Cognition, n.’ 2018. *OED Online*. Oxford University Press. www.oed.com/view/Entry/35876.

Columbus, Chris. 2000. *The Bicentennial Man*. Film. Columbia Pictures Corporation.

Comres. 2013. ‘Big Brother Watch Online Survey’. London: Comres.

Consumer Reports. 2016. ‘Tesla’s Autopilot: Too Much Autonomy Too Soon’. Consumer Reports. http://www.consumerreports.org/tesla/tesla-autopilot-too-much-autonomy-too-soon/.

Crampton, Eric P., and Donald J. Boudreaux. 2003. ‘Does Cyberspace Need Antitrust?’ In *Who Rules the Net?*, edited by Adam D Thierer and Clyde Wayne Jr. Crews. Washington, D.C.: Cato Institute.

Cranor, Lorrie, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. 2002. ‘The Platform for Privacy Preferences 1.0 (P3P1.0) Specification’. W3C Recommendation. W3C. http://www.w3.org/TR/P3P/.

Critical Appraisal Skills Programme (CASP). 2006. ‘10 Questions to Help You Make Sense of Qualitative Research’. Public Health Resource Unit.

Culnan, Mary J. 1993. ‘“How Did They Get My Name?”: An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use’. *MIS Quarterly* 17 (3): 341–63.

Curran, James, Natalie Fenton, and Des Freedman. 2012. *Misunderstanding the Internet*. London: Routledge.

Cusumano, Michael. 2010. ‘Technology Strategy and Management: The Evolution of Platform Thinking’. *Communications of the ACM* 53 (1): 32–34.

Dainow, Brandt. in-press. ‘Binding the Smart City Human-Digital System with Communicative Processes’. In *Technology and the City*, edited by Michael Nagenborg.

———. 2013. 'What Can a Medieval Friar Teach Us about the Internet: Deriving Criteria of Justice for Cyberlaw from Thomist Natural Law Theory'. *Philosophy and Technology* 26 (4): 459–76.

———. 2015a. 'Digital Alienation as the Foundation of Online Privacy Concerns'. *Computers & Society* ETHICOMP Special Issue: 109–17. https://doi.org/10.1145/2874239.2874255.

———. 2015b. 'Key Dialectics in Cloud Services'. *Computers & Society* ETHICOMP Special Issue: 52–59. https://doi.org/10.1145/2874239.2874247.

———. 2017a. 'Smart City Transcendent - Understanding the Smart City by Transcending Ontology'. *Orbit* 1. https://www.orbit-rri.org/volume-one/smart-city-transcendent/.

———. 2017b. 'Threats to Autonomy from Emerging ICTs'. *Australasian Journal of Information Systems* 21 (0). https://doi.org/10.3127/ajis.v21i0.1438.

Dator, Jim. 2011. 'Wendell Bell: The Futurist Who Would Put My Grandmother in Prison'. *Futures* 43 (6): 578–582.

Day, G.S., and P.J.H. Schoemaker. 2004. 'Driving through the Fog: Managing at the Edge'. *Long Range Planning* 37 (2): 127–42.

Deighton, John, and Lenora Kornfield. 2012. 'Economic Value of an Advertising-Supported Internet Ecosystem'. New York, N.Y: Internet Advertising Bureau. http://www.iab.net/insights_research/industry_data_and_landscape/economicvalue.

Dennis, Louise, Michael Fisher, Marija Slavkovik, and Matt Webster. 2013. 'Ethical Choice in Unforeseen Circumstances'. In *Towards Autonomous Robotic Systems*, 433–445. Springer.

Derry, Thomas, and Trevor Williams. 1993. *A Short History of Technology from the Earliest Times to AD 1900*. Oxford: Oxford Paperbacks.

Deslauriers, Marguerite. 2007. *Philosophia Antiqua, Volume 109 : Aristotle on Definition*. Boston, MA, USA: Brill Academic Publishers. http://site.ebrary.com/lib/nuim/docDetail.action?docID=10270883.

Di Paolo, Ezequiel A. 2005. 'Autopoiesis, Adaptivity, Teleology, Agency'. *Phenomenology and the Cognitive Sciences* 4 (4): 429–452.

Dillahunt, Tawanna R, Christopher A Brooks, and Samarth Gulati. 2015. 'Detecting and Visualizing Filter Bubbles in Google and Bing'. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 1851–1856. ACM.

Disco, Cornelis. 2005. 'Back to the Drawing Board: Inventing a Sociology of Technology'. *Inside the Politics of Technology: Agency and Normativity in the Co-Production of Technology and Society*, 29–60.

Dixon-Woods, Mary, Alex Sutton, Rachel Shaw, Tina Miller, Jonathan Smith, Bridget Young, Sheila Bonas, Andrew Booth, and David Jones. 2007. 'Appraising Qualitative Research for Inclusion in Systematic Reviews: A Quantitative and Qualitative Comparison of Three Methods'. *Journal of Health Services Research & Policy* 12 (1): 42–47. https://doi.org/10.1258/135581907779497486.

Dodge, Martin, and Rob Kitchin. 2005. 'Code and the Transduction of Space'. *Annals of the Association of American Geographers* 95 (1): 162–180.

Donchin, Anne. 2000. 'Autonomy and Interdependence: Quandaries in Genetic Decision-Making'. In *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, edited by Catriona Mackenzie and Natalie Stoljar, 236–58. New York: Oxford University Press.

Dong, Andy. 2004. 'Design as a Socio-Cultural Cognitive System'. In *DS 32: Proceedings of DESIGN 2004, the 8th International Design Conference, Dubrovnik, Croatia*.

Doorn, Jenny van, and JannyC. Hoekstra. 2013. 'Customization of Online Advertising: The Role of Intrusiveness'. *Marketing Letters* 24 (4): 339–51. https://doi.org/10.1007/s11002-012-9222-1.

Dreslinski, Ronald G, Michael Wieckowski, David Blaauw, Dennis Sylvester, and Trevor Mudge. 2010. 'Near-Threshold Computing: Reclaiming Moore's Law through Energy Efficient Integrated Circuits'. *Proceedings of the IEEE* 98 (2): 253–266.

Dryden, Jane. 2015. 'Autonomy'. *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/autonomy/.

Durling, David. 2013. 'Understanding the PhD by Publication'. In *Design Learning for Tomorrow: Design Education from Kindergarten to PhD.*, 3:1523–1533. http://durling.org/papers_files/cumulus%20submission%20v4.pdf.

Dworkin, Gerald. 1988. *The Theory and Practice of Autonomy*. Cambridge Studies in Philosophy. Cambridge; New York: Cambridge University Press.

———. 2015. 'The Nature of Autonomy'. *Nordic Journal of Studies in Educational Policy* 1 (0). https://doi.org/10.3402/nstep.v1.28479.

Dyhouse, Tony. 2010. 'Addressing the Silo Mentality'. *Infosecurity* 7 (2): 43. https://doi.org/10.1016/S1754-4548(10)70043-7.

Edwards, A., G. Elwyn, K. Hood, and S. Rollnick. 2000. 'Judging the "weight of Evidence" in Systematic Reviews: Introducing Rigour into the Qualitative Overview Stage by Assessing Signal and Noise.' *Journal of Evaluation in Clinical Practice* 6 (2): 177–84.

Ekstrom, Laura Waddell. 1993. 'A Coherence Theory of Autonomy'. *Philosophy and Phenomenological Research* 53 (3): 599–616.

Epstein, Robert, and Ronald E Robertson. 2015. 'The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections'. *Proceedings of the National Academy of Sciences* 112 (33): E4512–E4521.

Erlandsson, Fredrik, Martin Boldt, and Henric Johnson. 2012. 'Privacy Threats Related to User Profiling in Online Social Networks'. In *Proceedings of 2012 International Conference on Social Computing*, 838–842. Amsterdam: IEEE.

Eshraghi, Ali, and Stephen Harwood. 2013. 'Rethinking Affordances: Three Stories about the (Non-)Adoption of Digital Technologies'. In .

Ess, Charles. 2013. 'East-West Perspectives on Privacy, Ethical Pluralism and Global Information Ethics'. *From Ontos Verlag: Publications of the Austrian Ludwig Wittgenstein Society-New Series (Volumes 1-18)* 7.

'ETICA Project Website'. n.d. Accessed 25 July 2014. http://www.etica-project.eu/home.

European Parliament. 2011. 'Pathways towards Responsible ICT Innovation: Policy Brief of STOA on the ETICA Project'. Policy Brief PE 460.346. European Parliament: Directorate General For Internal Policies.

Facebook Inc. 2014. 'Facebook Inc. First Quarter 2014 Results'. Facebook Inc. http://investor.fb.com/releasedetail.cfm?ReleaseID=842071.

Federal Trade Commission. 2014. 'Data Brokers: A Call for Transparency'. Federal Trade Commission. https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

Filipponi, Luca, Andrea Vitaletti, Giada Landi, Vincenzo Memeo, Giorgio Laura, and Paolo Pucci. 2010. 'Smart City: An Event Driven Architecture for Monitoring Public Spaces with Heterogeneous Sensors'. In *Sensor Technologies and Applications (SENSORCOMM), 2010 Fourth International Conference On*, 281–286. IEEE.

Fisher, Franklin M. 2001. 'Innovative Industries and Antitrust: Comments on The Microsoft Antitrust Case'. *Journal of Industry, Competition and Trade* 1 (1): 41–52.

Floridi, Luciano. 2008. 'A Defence of Informational Structural Realism'. *Synthese* 161 (2): 219–253.

Fogel, Jeremy. 2014. 'A Reasonable Expectation of Privacy'. *Litigation* 40 (4). http://www.americanbar.org/publications/litigation_journal/2013-14/spring/a_reasonable_expectation_privacy.html.

Fontana, Roberto, and Lionel Nesta. 2009. 'Product Innovation and Survival in a High-Tech Industry'. *Review of Industrial Organization* 34 (4): 287–306. https://doi.org/10.1007/s11151-009-9210-7.

Frankfurt, Harry G. 1971. 'Freedom of the Will and the Concept of a Person'. *The Journal of Philosophy* 68 (1): 5–20.

Frick, Liezel. 2016. 'PHD by Publication: An Institutional Survey'. In *Postgraduate Supervision: Future Foci for the Knowledge Society*, edited by Magda Fourie-Malherbe, Ruth Albertyn, and Eli Bitzer, 299–312. Stellenbosch, South Africa: Sun Press.

Fuchs, Christian. 2013. 'Class and Exploitation on the Internet'. In *Digital Labor - the Internet as Playground and Factory*, edited by Trebor Scholz, 211–24. New York: Routledge.

———. 2014. *Digital Labour and Karl Marx*. New York, NY: Routledge, Taylor & Francis Group.

Fuchs, Christian, and Marisol Sandoval. 2014. 'Digital Workers of the World Unite! A Framework for Critically Theorising and Analysing Digital Labour'. *TripleC* 12 (2): 486–563.

Galič, Maša, Tjerk Timan, and Bert-Jaap Koops. 2017. 'Bentham, Deleuze and Beyond: An Overview of Surveillance Theories from the Panopticon to Participation'. *Philosophy & Technology* 30 (1): 9–37. https://doi.org/10.1007/s13347-016-0219-1.

Galitz, Wilbert O. 2007. *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques*. New York, N.Y: John Wiley & Sons.

Gates, Arlan. 2000. 'Convergence and Competition: Technological Change, Industry Concentration and Competition Policy in the Telecommunications Sector'. *U. Toronto Fac. L. Rev.* 58: 83.

Georghiou, Luke, and Michael Keenan. 2006. 'Evaluation of National Foresight Activities: Assessing Rationale, Process and Impact'. *Technological Forecasting and Social Change* 73 (7): 761–77. https://doi.org/10.1016/j.techfore.2005.08.003.

Gershon, Ilana. 2011. 'Un-Friend My Heart: Facebook, Promiscuity, and Heartbreak in a Neoliberal Age'. *Anthropological Quarterly* 84 (4): 865–94.

Gibson, James Jerome. 1986. *The Ecological Approach to Visual Perception*. New York: Psychology Press.

Glenn, Jerome C. 2009a. 'Genius Forecasting, Intuition, and Vision'. In *Futures Research Methods*, 3rd ed. Washington, DC: The Millenium Project.

———. 2009b. 'Introduction to Futures Research Methods'. In *Futures Research Methods*, edited by Jerome C Glenn and Theodore J Gordon, 3rd ed. Washington, DC: The Millenium Project.

———. 2009c. 'Scenarios'. In *Futures Research Methods*, edited by Jerome C Glenn and Theodore J Gordon, 3rd ed. Washington, D.C.: The Millenium Project.

Glenn, Jerome C, and Theodore J Gordon, eds. 2009. *Futures Research Methods*. 3rd ed. Washington, DC: The Millennium Project.

Godet, Michel. 2009. 'Structural Analysis'. In *Futures Research Methods*, 3rd ed. Washington, DC: The Millenium Project.

Goffman, Erving. 1990. *The Presentation of Self in Everyday Life*. Nachdr. Anchor Books. New York, NY: Doubleday.

Goldsmith, Mike. 2012. *Discord: The Story of Noise*. Oxford: Oxford University Press.

Google Inc. 2014. 'Google Inc. First Quarter 2014 Results'. Google Inc. http://investor.google.com/earnings/2014/Q1_google_earnings.html.

Gordon, Theodore J, and Jerome C Glenn. 2004. "Integration, Comparisons, and Frontiers of Futures Research Methods'. In *EU-US Seminar: New Technology Foresight, Forecasting & Assessment Methods*. Vol. 1314.

———. 2009. 'Environmental Scanning'. In *Futures Research Methods*. Washington, DC: The Millenium Project.

Goujon, Philippe, and Catherine Flick. 2011. 'D.4.1 Governance Approaches'. ETICA Project.

Graham, Gordon. 2004. *Eight Theories of Ethics*. London; New York: Routledge/Taylor and Francis Group.

Graziani, R., and A. Johnson. 2008. *Routing Protocols and Concepts: CCNA Exploration Companion Guide*. Indianapolis: Cisco Systems, Inc.

Grunwald, Armin. 1999. 'Technology Assessment or Ethics of Technology? Reflections on Technology Development between Social Sciences and Philosophy.' *Ethical Perspect* 6 (2): 171 – 182.

Gubbi, Jayavardhana, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. 2013. 'Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions'. *Future Generation Computer Systems* 29 (7): 1645–1660.

Guo, Bin, Zhu Wang, Daqing Zhang, Zhiwen Yu, and Xingshe Zhou. 2013. 'Opportunistic IoT: Exploring the Harmonious Interaction between Human and the Internet of Things'. *Journal of Network and Computer Applications* 36 (6): 1531–1539.

Guyer, Paul. 2003. 'Kant on the Theory and Practice of Autonomy'. In *Autonomy*, edited by Ellen Frankel Paul, Fred Dycus Miller, and Jeffrey Paul, 70–98. Cambridge: Cambridge University Press.

Hacklin, Fredrik, Christian Marxt, and Fritz Fahrni. 2009. 'Coevolutionary Cycles of Convergence: An Extrapolation from the ICT Industry'. *Technological Forecasting and Social Change* 76 (6): 723–36. https://doi.org/10.1016/j.techfore.2009.03.003.

Hadley, Herbert Spencer. 1895. 'Right to Privacy'. *NWL Rev.* 3: 1.

Haegeman, Karel, Elisabetta Marinelli, Fabiana Scapolo, Andrea Ricci, and Alexander Sokolov. 2013. 'Quantitative and Qualitative Approaches in Future-Oriented Technology Analysis (FTA): From Combination to Integration?' *Technological Forecasting and Social Change* 80 (3): 386–397.

Halliday, Tim, ed. 1998. *Biology: Brain and Behaviour—The Senses and Communication*. Heidelberg: Springer-Verlag.

Hamburger, Tom, and Matea Gold. 2014. 'Google, Once Disdainful of Lobbying, Now a Master of Washington Influence'. *The Washington Post*, 12 April 2014. http://www.washingtonpost.com/politics/how-google-is-transforming-power-and-politicsgoogle-once-disdainful-of-lobbying-now-a-master-of-washington-influence/2014/04/12/51648b92-b4d3-11e3-8cb6-284052554d74_story.html.

Hancke, Gerhard P., Bruno de Carvalho e Silva, and Gerhard P. Hancke Jr. 2013. 'The Role of Advanced Sensing in Smart Cities'. *Sensors* 13 (1): 393. https://doi.org/10.3390/s130100393.

Hannaford, Steve. 2007. *Market Domination!: The Impact of Industry Consolidation on Competition, Innovation, and Consumer Choice*. Westport, CT.: Praeger Publishers.

Hannes, K. 2011. 'Chapter 4: Critical Appraisal of Qualitative Research'. In *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions*, edited by K Noyes, A Booth, K Hannes, A. Harden, J Harris, S

Lewin, and C Lockwood. Cochrane Collaboration Qualitative Methods Group. http://cqrmg.cochrane.org/supplemental-handbook-guidance.

Harris, Ian, Penny Duquenoy, Richard Jennings, David Pullinger, and Simon Rogerson. 2010. 'Ethical Assessment of New Technologies: A Meta-Methodology'. *Journal of Information, Communication and Ethics in Society* 9 (1): 49–64.

Hartson, Rex. 2003. 'Cognitive, Physical, Sensory, and Functional Affordances in Interaction Design'. *Behaviour & Information Technology* 22 (5): 315–338.

Hartswood, Mark, Barbara Grimpe, Marina Jirotka, and Stuart Anderson. 2014. 'Towards the Ethical Governance of Smart Society'. In *Social Collective Intelligence*, edited by Daniele Miorandi, Vincenzo Maltese, Michael Rovatsos, Anton Nijholt, and James Stewart, 3–30. London; New York: Springer.

Heersmink, Richard, Jeroen den Hoven, Nees Jan Eck, and Jan den Berg. 2011. 'Bibliometric Mapping of Computer and Information Ethics'. *Ethics and Information Technology* 13 (3): 241–49. https://doi.org/10.1007/s10676-011-9273-7.

Heersmink, Richard, Jeroen van den Hoven, and Job Timmermans. 2010. 'D.2.2 Normative Issues Report'. D.2.2. ETICA Project.

Heidegger, Martin. 1977. *The Question Concerning Technology, and Other Essays*. New York: Harper & Row.

Hildebrandt, M. 2008. 'Defining Profiling: A New Type of Knowledge?' In *Profiling the European Citizen*, edited by M Hildebrandt and Serge Gurtwith. New York, NY: Springer.

Hillis, Ken, Kylie Jarrett, and Michael Petit. 2013. *Google and the Culture of Search*. New York: Routledge.

Hocutt, Max. 1974. 'Aristotle's Four Becauses'. *Philosophy* 49 (190): 385–399.

Hof, Simone van der, and Corien Prins. 2008. 'Personalisation and Its Influence on Identities, Behaviour and Social Values'. In *Profiling the European Citizen*, edited by M Hildebrandt and Serge Gurtwith, 111–27. New York, N.Y: Springer.

Hull, Gordon, Heather Richter Lipford, and Celine Latulipe. 2011. 'Contextual Gaps: Privacy Issues on Facebook'. *Ethics and Information Technology* 13 (4): 289–302. https://doi.org/10.1007/s10676-010-9224-8.

Huston, Geoff. 1999. *ISP Survival Guide: Strategies for Running a Competitive ISP*. New York, N.Y: Wiley.

Hutchby, Ian. 2001. 'Technologies, Texts and Affordances'. *Sociology* 35 (2): 441–56. https://doi.org/10.1177/S0038038501000219.

———. 2003. 'Affordances and the Analysis of Technologically Mediated Interaction: A Response to Brian Rappert'. *Sociology* 37 (3): 581–89. https://doi.org/10.1177/00380385030373011.

Huxley, Aldous. 2006. *Brave New World*. London: Harper Perennial Modern Classics.

Ide, Nancy. 2007. 'Chapter 21. Preparation and Analysis of Linguistic Corpora'. In *A Companion to Digital Humanities*, edited by S. Schreibman, R. Siemens, and J. Unsworth. Malden, MA: Blackwell Publishing Ltd. http://onlinelibrary.wiley.com/doi/10.1002/9780470999875.ch21/summary.

Ikonen, Veikko, Minni Kanerva, Panu Kouri, Bernd Stahl, and Kutoma Wakunuma. 2010. 'D.1.2. Emerging Technologies Report'. D.1.2. ETICA Project.

Introna, Lucas. 2011. 'Phenomenological Approaches to Ethics and Information Technology'. Stanford Encyclopedia of Philosophy. 2011. http://plato.stanford.edu/entries/ethics-it-phenomenology/.

ITU-T. 2008. 'Ubiquitous Sensor Networks (USN)'. 4. ITU-T Technology Watch Briefing Report Series. International Telecommunications Union.

Jaisingh, Kushal, Khalil El-Khatib, and Rajen Akalu. 2016. 'Paving the Way for Intelligent Transport Systems (ITS): Privacy Implications of Vehicle Infotainment and Telematics Systems'. In *Proceedings of the 6th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications*, 25–31. DIVANet '16. New York, NY, USA: ACM. https://doi.org/10.1145/2989275.2989283.

Jiang, Hao, and John M Carroll. 2009. 'Social Capital, Social Network and Identity Bonds: A Reconceptualization'. In *Proceedings of the Fourth International Conference on Communities and Technologies*, 51–60. ACM.

Jin, Jiong, Jayavardhana Gubbi, Slaven Marusic, and Marimuthu Palaniswami. 2014. 'An Information Framework for Creating a Smart City through Internet of Things'. *IEEE Internet of Things Journal* 1 (2): 112–121.

Johnson, Deborah G. 2011. 'Software Agents, Anticipatory Ethics, and Accountability'. In *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight*, edited by Gary E. Marchant, Braden R. Allenby, and Joseph R. Herkert, 7:61–76. Dordrecht: Springer Netherlands. http://www.springerlink.com/index/10.1007/978-94-007-1356-7_5.

Johnson, Eric J, Suzanne B Shu, Benedict GC Dellaert, Craig Fox, Daniel G Goldstein, Gerald Häubl, Richard P Larrick, John W Payne, Ellen Peters, and David Schkade. 2012. 'Beyond Nudges: Tools of a Choice Architecture'. *Marketing Letters* 23 (2): 487–504.

Johnson, Robert. 2014. 'Kant's Moral Philosophy'. Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/sum2014/entries/kant-moral/.

Jonas, Hans. 1984. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago: University of Chicago Press.

Jung, C. G. 1977. *The Collected Works of C. G. Jung. 6: Psychological Types*. Edited by Herbert Read. Bollingen Series. New York, NY: Pantheon Books.

Kahn, H., and A.J. Wiener. 1967. *The Year 2000: A Framework for Speculation on the Next Thirty-Three Years*. New York: Macmillan.

Kant, Immanuel. 1785. *Grundlegung Zur Metaphysik Der Sitten*. Riga: Johann Friedrich Harttnoch.

———. 1998. *Groundwork of the Metaphysics of Morals*. Translated by Mary J. Gregor. Cambridge Texts in the History of Philosophy. New York: Cambridge University Press.

Karger, Paul A., and Andrew J. Herbert. 1984. 'An Augmented Capability Architecture to Support Lattice Security and Traceability of Access'. In , 2–2. IEEE. https://doi.org/10.1109/SP.1984.10001.

Kellner, Douglas. 1990. 'From 1984 to One-Dimensional Man: Critical Reflections on Orwell and Marcuse'. *Current Perspectives in Social Theory* 10: 223–52.

Kennedy, Helen. 2008. 'New Media's Potential for Personalization'. *Information, Communication & Society* 11 (3): 307–25. https://doi.org/10.1080/13691180802025293.

Killeen, Peter R, and Kenneth W Jacobs. 2016. 'Coal Is Not Black, Snow Is Not White, Food Is Not a Reinforcer: The Roles of Affordances and Dispositions in the Analysis of Behavior'. *The Behavior Analyst* 40 (1): 17–38.

Kim, Hyun Chan, Seongcheol Mun, Hyun-U Ko, Lindong Zhai, Abdullahil Kafy, and Jaehwan Kim. 2016. 'Renewable Smart Materials'. *Smart Materials and Structures* 25 (7): 073001.

Kim, R., and S. Kaplan. 2005. 'Co-Evolution in Information Systems Engagement: Exploration, Ambiguity and the Emergence of Order". In *3rd Int. Conf. on Action in Language, Organisations and Information Systems*, 166–180. Limerick, Ireland.

Kim, Y, Marie Kim, and Yong Joon Lee. 2008. 'COSMOS: A Middleware Platform for Sensor Networks and a U-Healthcare Service'. In *Proceedings of the 2008 ACM Symposium on Applied Computing*, 512–513. SAC '08. New York, NY, USA: ACM. https://doi.org/10.1145/1363686.1363812.

Kitchin, R. 2016. 'Getting Smarter about Smart Cities: Improving Data Privacy and Data Security'. Dublin, Ireland: Data Protection Unit, Department of the Taoiseach.

Koh, Byungwan. 2011. *User Profiling in Online Marketplaces and Security*. Ann Arbor, MI: ProQuest LLC.

Komninos, Nicos. 2006. 'The Architecture of Intelligent Cities: Integrating Human, Collective and Artificial Intelligence to Enhance Knowledge and Innovation'. In *Intelligent Environments, 2006. IE 06. 2nd IET International Conference On*, 1:13–20. IET.

Komninos, Nicos, Hans Schaffers, and Marc Pallot. 2011. 'Developing a Policy Roadmap for Smart Cities and the Future Internet'. In *EChallenges E-2011 Conference Proceedings, IIMC International Information Management Corporation*. IMC International Information Management Corporation.

Koops, Bert-Jaap, and Ronald Leenes. 2014. 'Privacy Regulation Cannot Be Hardcoded. A Critical Comment on the "Privacy by Design" Provision in Data-Protection Law'. *International Review of Law, Computers & Technology* 28 (2): 159–71. https://doi.org/10.1080/13600869.2013.801589.

Korolova, Aleksandra. 2011. 'Privacy Violations Using Microtargeted Ads: A Case Study'. *Journal of Privacy and Confidentiality* 3 (1).

Kothari, Chakravanti Rajagopalachari. 2004. *Research Methodology: Methods and Techniques*. New Age International.

Kramer, A. D. I., J. E. Guillory, and J. T. Hancock. 2014. 'Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks'. *Proceedings of the National Academy of Sciences* 111 (24): 8788–90. https://doi.org/10.1073/pnas.1320040111.

Krämer, Nicole C., and Stephan Winter. 2008. 'Impression Management 2.0: The Relationship of Self-Esteem, Extraversion, Self-Efficacy, and Self-Presentation Within Social Networking Sites'. *Journal of Media Psychology* 20 (3): 106–16. https://doi.org/10.1027/1864-1105.20.3.106.

Krasnova, Hanna, Oliver Günther, Sarah Spiekermann, and Ksenia Koroleva. 2009. 'Privacy Concerns and Identity in Online Social Networks'. *Identity in the Information Society* 2 (1): 39–63. https://doi.org/10.1007/s12394-009-0019-1.

Kraut, Richard. 2016. 'Aristotle's Ethics'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2016. http://plato.stanford.edu/archives/spr2016/entries/aristotle-ethics/.

Krefting, Laura. 1991. 'Rigor in Qualitative Research: The Assessment of Trustworthiness'. *American Journal of Occupational Therapy* 45 (3): 214–22. https://doi.org/10.5014/ajot.45.3.214.

Kroes, Neelie, and Carl-Christian Buhr. 2014. 'Human Society in a Digital World'. Digital Minds for a New Europe. 2014. https://ec.europa.eu/commission_2010-2014/kroes/en/content/human-society-digital-world-neelie-kroes-and-carl-christian-buhr.

Krüger, Steffan, and Jacob Johanssen. 2014. 'Alienation and Digital Labour—A Depth-Hermeneutic Inquiry into Online Commodification and the Unconscious'. *TripleC* 12 (2): 632–45.

Kühler, Michael, and Nadja Jelinek, eds. 2013. *Autonomy and the Self.* Dordrecht: Springer Netherlands. http://link.springer.com/10.1007/978-94-007-4789-0.

Lair, D. J. 2005. 'Marketization and the Recasting of the Professional Self: The Rhetoric and Ethics of Personal Branding'. *Management Communication Quarterly* 18 (3): 307–43. https://doi.org/10.1177/0893318904270744.

Lakatos, Imre. 1970. 'Falsification and the Methodology of Scientific Research Programmes'. In *Criticism and the Growth of Knowledge*, edited by Imre Lakatos and Alan Musgrave, 170–96. Cambridge, UK: Cambridge University Press.

Lamb, Roberta, and Rob Kling. 2003. 'Reconceptualizing Users as Social Actors in Information Systems Research'. *MIS Quarterly* 27 (2): 197–236.

Langheinrich, Marc. 2001. 'Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems'. In *Proceedings of the Third International Conference on Ubiquitous Computing*, edited by Gregory Abowd, Barry Brumitt, and Steven Shafer, 273–91. LNCS. Atlanta, USA: Springer-Verlag. http://www.vs.inf.ethz.ch/publ/papers/privacy-principles.pdf.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Clarendon Lectures in Management Studies. Oxford; New York: Oxford University Press.

———. 2011. 'Network Theory| Networks, Societies, Spheres: Reflections of an Actor-Network Theorist'. *International Journal of Communication* 5: 15.

Law, John. 1992. 'Notes on the Theory of the Actor-Network: Ordering, Strategy, and Heterogeneity'. *Systemic Practice and Action Research* 5 (4): 379–393.

Lazarus, David. 2015. 'Verizon's Super New Way to Mess with Your Privacy'. *The Los Angeles Times*, 3 February 2015.

Lightcap, Tracey. 2011. 'Review: Democracy, Law and Governance by Jacque Lenoble and Marc Maesschalck'. *Law & Politics Book Review* 27 (9): 490–94.

Limonard, Sander, and Nicole de Koning. 2005. 'Dealing with Dilemmas in Pre-Competitive ICT Development Projects: The Construction of "The Social" in Designing New Technologies'. In *Everyday Innovators*, edited by Leslie Haddon, Enid Mante, Bartolomeo Sapio, Kari-Hans Kommonen, Leopoldina Fortunati, and Annevi Kant, 32:168–83. Computer Supported Cooperative Work. Springer Netherlands. http://dx.doi.org/10.1007/1-4020-3872-0_11.

Liu, Yabing, Krishna P Gummadi, Balachander Krishnamurthy, and Alan Mislove. 2011. 'Analyzing Facebook Privacy Settings: User Expectations vs. Reality'. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, 61–70. ACM.

Lo, Benny, Surapa Thiemjarus, Rachel King, and Guang-Zhong Yang. 2005. 'Body Sensor Network–a Wireless Sensor Platform for Pervasive Healthcare Monitoring'. In *Adjunct Proceedings of the 3rd International Conference on Pervasive Computing*. Hagenberg, Austria: Springer.

Luhmann, Niklas. 1986a. 'The Autopoiesis of Social Systems'. *Sociocybernetic Paradoxes*, 172–192.

———. 1986b. 'The Autopoiesis of Social Systems'. In *Sociocybernetic Paradoxes: Observation, Control and Evolution of Self-Steering Systems*, edited by Felix Geyer and Johannes van der Zouwen, 172–192. London: SAGE Publications.

———. 1995. *Social Systems*. Stanford, Calif: Stanford University Press.

Lyon, David, Kirstie Ball, and Kevin D Haggerty. 2012. *Routledge Handbook of Surveillance Studies*. Abingdon, Oxon; New York, NY: Routledge.

MacIntyre, Alan. 1998. *A Short History of Ethics*. 2nd ed. Padstow: Routledge.

MacMillian, Neil A. 2002. 'Signal Detection Theory'. In *Stevens' Handbook of Experimental Psychology*, edited by John Wikted, 3rd ed., 4:43–90. New York, N.Y.: John Wiley & Sons, Inc.

Madejski, Michelle, Maritza Johnson, and Steven M Bellovin. 2012. 'A Study of Privacy Settings Errors in an Online Social Network'. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference On*, 340–345. IEEE.

Manders-Huits, Noëmi. 2011. 'What Values in Design? The Challenge of Incorporating Moral Values into Design'. *Science and Engineering Ethics* 17 (2): 271–87. https://doi.org/10.1007/s11948-010-9198-2.

Mansell, Robin. 2017. 'Imaginaries of the Digital: Ambiguity, Power and the Question of Agency'. *Communiquer. Revue de Communication Sociale et Publique*, no. 20 (September): 40–48. https://doi.org/10.4000/communiquer.2261.

Marshall, Jill. 2008. *Personal Freedom through Human Rights Law*. Leiden; Boston: Martinus Nijhoff.

Marthews, Alex, and Catherine Tucker. 2014. 'Government Surveillance and Internet Search Behavior'. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2412564.

Martin, Silva, and Stef Nicovich. 2013. 'Self Concept Clarity and Its Impact on the Self-Avatar Relationship in a Mediated Environment'. In *Atlantic Marketing Association Conference Proceedings*. Atlanta, USA.

Martínez-Ballesté, Antoni, Pablo A Pérez-Martínez, and Agusti Solanas. 2013. 'The Pursuit of Citizens' Privacy: A Privacy-Aware Smart City Is Possible'. *IEEE Communications Magazine* 51 (6): 136–141.

Mastrogiovanni, Fulvio, Antonio Sgorbissa, and Renato Zaccaria. 2010. 'From Autonomous Robots to Artificial Ecosystems'. In *Handbook of Ambient Intelligence and Smart Environments*, edited by Hideyuki Nakashima, Juan Carlos Augusto, and Hamid Aghajan, 635–668. Springer.

Matthias, Andreas. 2004. 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata'. *Ethics and Information Technology* 6 (3): 175–83. https://doi.org/10.1007/s10676-004-3422-1.

Mattli, Walter. 2001. 'The Politics and Economics of International Institutional Standards Setting: An Introduction'. *Journal of European Public Policy* 8 (3): 328–344.

Maturana, Humberto R. 1981. 'The Organization of the Living: A Theory of the Living Organization'. *Cybernetics Forum* X (2–3): 14–23.

Mayer-Schönberger, Viktor. 2009. *Delete: The Virtue of Forgetting in the Digital Age*. Princeton, NJ: Princeton Univ. Press.

McChesney, Robert W. 2014. 'Be Realistic, Demand the Impossible: Three Radically Democratic Internet Policies'. *Critical Studies in Media Communication* 31 (2): 92–99. https://doi.org/10.1080/15295036.2014.913806.

McCoy, Damon, Kevin Bauer, Dirk Grunwald, Tadayoshi Kohno, and Douglas Sicker. 2008. 'Shining Light in Dark Places: Understanding the Tor Network'. In *Privacy Enhancing Technologies*, 63–76. Springer.

McDonald, Aleecia M, and Lorrie Faith Cranor. 2008. 'The Cost of Reading Privacy Policies'. *ISJLP* 4: 543.

McKay, Susan R, Wolfgang Christian, James R Matey, and Edward Tufte. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphics Press.

Meeder, Brendan, Jennifer Tam, Patrick Gage Kelley, and Lorrie Faith Cranor. 2010. 'RT@ IWantPrivacy: Widespread Violation of Privacy Settings in the Twitter Social Network'. In *Proceedings of the Web*, 2:1–12. Pittsburgh, USA.

Mehdizadeh, Soraya. 2010. 'Self-Presentation 2.0: Narcissism and Self-Esteem on Facebook'. *Cyberpsychology, Behavior, and Social Networking* 13 (4): 357–64. https://doi.org/10.1089/cyber.2009.0257.

Mell, Peter, and Timothy Grance. 2011. 'The NIST Definition of Cloud Computing'. Special Publication 800-145. National Institude for Standards & Technology.

Meyers, Diana T. 1989a. 'Part 2.2 An Alternative Account of Autonomy'. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1321773.

———. 1989b. *Self, Society, and Personal Choice*. New York: Columbia University Press.

———. 2005. 'Decentralizing Autonomy: Five Faces of Selfhood'. In *Autonomy and the Challenges to Liberalism*, edited by John Christman and Joel Anderson, 27–55. Cambridge: Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9780511610325A011.

Mierzwinski, Edmund, and Jeffrey Chester. 2013. 'Selling Consumers Not Lists: The New World of Digital Decision-Making and the Role of the Fair Credit Reporting Act'. *Suffolk University Law Review* 46 (3): 855–67.

Miller, Bruce L. 1981. 'Autonomy & the Refusal of Lifesaving Treatment'. *Hastings Center Report* 11 (4): 22–28.

Miller, Hugh, and Jill Arnold. 2001. 'Self in Web Home Pages'. In *Towards CyberPsychology*, edited by Giuseppe Riva and Carlo Galimberti, 73–94. Studies in New Technologies and Practices in Communication. Oxford: IOS Press.

Miller, Seumas. 2008. 'Collective Responsibility and Information and Communication Technology'. In *Information Technology and Moral Philosophy*, edited by Jeroen van den Hoven and John Weckert, 226–50. Cambridge; New York: Cambridge University Press.

Misa, Thomas J. 1994. 'Retrieving Sociotechnical Change from Technological Determinism'. In *Does Technology Drive History?: The Dilemma of Technological Determinism*, edited by Merritt Roe Smith and Leo Marx. Cambridge, Mass.: MIT Press.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Tadeo, Luciano Floridi, and Sandra Watchter. 2016. 'The Ethics of Algorithms: Mapping the Debate'. *Big Data & Society* 3 (2).

Moglen, Eben. 2010. 'Freedom in the Cloud'. Speech, New York, N.Y. http://www.softwarefreedom.org/events/2010/isoc-ny/FreedomInTheCloud-transcript.html.

Mollick, Ethan. 2006. 'Establishing Moore's Law'. *IEEE Annals of the History of Computing* 28 (3): 62–75.

Mueller, Milton. 1999. 'Digital Convergence and Its Consequences'. *Javnost-the Public* 6 (3): 11–27.

Musk, Elon. 2016. 'Tesla Master Plan, Part Deux'. *Personal Blog* (blog). 20 July 2016. https://www.tesla.com/blog/master-plan-part-deux.

Nagulendra, Sayooran, and Julita Vassileva. 2014. 'Understanding and Controlling the Filter Bubble through Interactive Visualization: A User Study'. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, 107–15. Santiago, Chile: ACM Press. https://doi.org/10.1145/2631775.2631811.

Nam, Taewoo, and Theresa A Pardo. 2011. 'Conceptualizing Smart City with Dimensions of Technology, People, and Institutions'. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, 282–291. ACM.

National Conference of State Legislatures. 2017. 'Autonomous | Self-Driving Vehicles Legislation'. National Conference of State Legislatures.

http://www.ncsl.org/research/transportation/autonomous-vehicles-legislation.aspx.

Nawyn, Jason, Stephen S Intille, and Kent Larson. 2006. 'Embedding Behavior Modification Strategies into a Consumer Electronic Device: A Case Study'. In *International Conference on Ubiquitous Computing*, 297–314. Springer.

Negroponte, Nicholas. 1996. *Being Digital.* 1. Vintage Books ed. New York, NY: Vintage Books.

Ng, Irene CL, and Susan YL Wakenshaw. 2015. 'Engineering a Platform for Personal Data as a Service: The Economic Model for the HAT (Hub of All Things)'. In *The 2015 Naples Forum on Service Service: Dominant Logic, Network and Systems Theory and Service Science.*

NICE. 2007. 'Appendix H: Methodology Checklist: Qualitative Studies'. In *The Guidelines Manual.* National Institute for Health and Clinical Excellence.

———. 2012. 'Appendix B: Methodology Checklist: Systematic Reviews and Meta-Analyses'. In *Appendices B–I.* The Guidelines Manual. National Institute for Health and Care Excellence. http://publications.nice.org.uk/PMG6B.

Niebuhr, R. 2004. *The Nature and Destiny of Man: A Christian Interpretation: Human Nature.* Library of Theological Ethics. Westminster, John Knox Press.

Nissenbaum, Helen Fay. 2001. 'How Computer Systems Embody Values'. *Computer* 34 (3): 118–20. https://doi.org/10.1109/2.910905.

Noam, Eli M, Donald Hay, Michael R Baye, and John Morgan. 2003. 'The Internet: Still Wide Open and Competitive?' *OII Internet Issue Brief*, no. 1.

Norman, Donald A. 2002. *The Design of Everyday Things.* New York: Basic Books.

Northcote, Maria T. 2012. 'Selecting Criteria to Evaluate Qualitative Research'. In *Narratives of Transition: Perspectives of Research Leaders, Educators & Postgraduates. Proceedings from the 10th Quality in Postgraduate Research Conference*, 99–110. Adelaide: The Centre for Higher Education, Learning and Teaching. The Australian National University. http://research.avondale.edu.au/edu_papers/38.

OECD. 2013a. 'OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data'. C(80)58/FINAL. OECD. http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm.

———. 2013b. 'Exploring the Economics of Personal Data'. OECD Digital Economy Papers 220. http://www.oecd-ilibrary.org/science-and-technology/exploring-the-economics-of-personal-data_5k486qtxldmq-en.

Office of the Data Protection Commissioner. 2015. 'Annual Report of the Data Protection Commissioner of Ireland 2014'. Data Protection Commission of Ireland.

Office of the Information & Privacy Commissioner of Ontario, and IBM. 2011. 'Privacy by Design: From Theory to Practice'. Ontario: Office of the Information & Privacy Commissioner of Ontario.

Ogilvy, James. 2002. 'Futures Studies and the Human Sciences: The Case for Normative Scenarios'. In *New Thinking for a New Millennium: The Knowledge Base of Futures Studies*, edited by Richard Slaughter. London: Routledge.

Ollman, Bertell. 2001. *Alienation: Marx's Concept of Man in Capitalist Society.* 2nd ed. Cambridge; New York: Cambridge University Press.

Olson, Eric T. 2011. 'The Extended Self'. *Minds and Machines* 21 (4): 481–495.

O'Neil, Cathy. 2017. *Weapons of Math Destruction.* New York, N.Y: Broadway Books.

Oppliger, Rolf, ed. 2014. *Secure Messaging on the Internet.* Boston: Artech House.

Orwell, G. 1983. *1984*. London: Houghton Mifflin Harcourt.

OSI Project. 1994. 'Information Technology – Open Systems Interconnection (OSI) – Basic Reference Model: The Basic Model'. Recommendation 200 (1994)| ISO/IEC 7498-1: 1994. International Organization for Standardization.

Palem, Krishna V, Lakshmi NB Chakrapani, Zvi M Kedem, Avinash Lingamneni, and Kirthi Krishna Muntimadugu. 2009. 'Sustaining Moore's Law in Embedded Computing through Probabilistic and Approximate Design: Retrospects and Prospects'. In *Proceedings of the 2009 International Conference on Compilers, Architecture, and Synthesis for Embedded Systems*, 1–10. ACM.

Palmås, Karl. 2011. 'Panspectric Surveillance and the Contemporary Corporation'. *Surveillance & Society* 8 (3): 338–54.

Parsons, Talcott. 1991. *The Social System*. London: Routledge.

Paul, Ellen Frankel, Fred Dycus Miller, and Jeffrey Paul, eds. 2003. *Autonomy*. Cambridge: Cambridge University Press.

Pellé, Sophie, and Bernard Reber. 2013. 'Governance of Responsible Innovation - Theoretical Landscape'. DEL.2.2. GREAT Deliverables. Governance for Responsible Innovation (GREAT). http://www.great-project.eu/deliverables_files/deliverables03.

Perera, Charith, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. 2014. 'Sensing as a Service Model for Smart Cities Supported by Internet of Things'. *Transactions on Emerging Telecommunications Technologies* 25 (1): 81–93.

Petrolo, Riccardo, Valeria Loscrì, and Nathalie Mitton. 2015. 'Towards a Smart City Based on Cloud of Things, a Survey on the Smart City Vision and Paradigms'. *Transactions on Emerging Telecommunications Technologies*. https://doi.org/10.1002/ett.2931.

Pettersson, Jan, Mirko Presser, Jesús Bernat Vercher, Luis Muñoz, José A Galache, Luis A Hernández Gómez, and José M Hernández-Muñoz. 2011. 'Smart Cities at the Forefront of the Future Internet'. In *The Future Internet Assembly*, 447–462. Springer.

Picard, Rosalind W. 2003. 'Affective Computing: Challenges'. *International Journal of Human-Computer Studies* 59 (1): 55–64.

Picon, Antoine. 2015. *Smart Cities: A Spatialised Intelligence*. Chichester; Malden, MA: John Wiley & Sons.

Pil Choi, Jay, and Byung-Cheol Kim. 2010. 'Net Neutrality and Investment Incentives'. *The RAND Journal of Economics* 41 (3): 446–471.

Poli, Roberto. 2011. 'Ethics and Future Studies'. *Int. J. Management Concepts and Philosophy* 5 (4): 403–10.

Popper, Rafael. 2009. *Mapping Foresight: Revealing How Europe and Other World Regions Navigate into the Future*. Luxembourg: European Union.

Porter, Alan L. 2009. 'Text Mining of Science & Technology Information Resources for Future-Oriented Technology Analyses'. In *Futures Research Methods*, 3rd ed. Washington, DC: The Millenium Project.

Poteralska, Beata, and Anna Sacio-Szymańska. 2014. 'Evaluation of Technology Foresight Projects'. *European Journal of Futures Research* 2 (1). https://doi.org/10.1007/s40309-013-0026-1.

Pouwelse, Johan, Paweł Garbacki, Dick Epema, and Henk Sips. 2005. 'The Bittorrent P2P File-Sharing System: Measurements and Analysis'. In *Peer-to-Peer Systems IV*, edited by Miguel Castro and Robbert van Renesse, 205–16. Lecture Notes in Computer Science 3640. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/11558989_19.

Powell, James R. 2008. 'The Quantum Limit to Moore's Law'. *Proceedings of the IEEE* 96 (8): 1247–1248.

‘Privacy | Samsung UK’. n.d. Accessed 18 June 2015. http://www.samsung.com/uk/info/privacy-SmartTV.html.

Psyllidis, Achilleas. 2015. ‘Ontology-Based Data Integration from Heterogeneous Urban Systems: A Knowledge Representation Framework for Smart Cities’. In *CUPUM 2014: Proceedings of the 14th International Conference on Computers in Urban Planning and Urban Management 2015*. Cambridge, Mass.: MIT.

Pursell, Carroll. 2007. *The Machine in America: A Social History of Technology*. Baltimore: John Hopkins University Press.

Rachlin, Howard, and Bryan A. Jones. 2010. ‘The Extended Self.’ In *Impulsivity: The Behavioral and Neurological Science of Discounting.*, edited by Gregory J. Madden and Warren K. Bickel, 411–31. Washington: American Psychological Association. http://content.apa.org/books/12069-015.

Rader, Michael, A. Antener, Rafael Capurro, Michael Nagenborg, and L. Stengel. 2010. ‘D.3.2 Evaluation Report’. D.3.2. ETICA Project.

Rainey, Stephen, and Philippe Goujon. 2011. ‘D.4.2 Governance Recommendations’. D.4.2. ETICA Project. http://www.academia.edu/download/30848560/Deliverable4.2.pdf.

Rapaport, William J. 2017. ‘What Is Computer Science’. *University at Buffalo, The State University*.

RDF Working Group. 2015. ‘RDF - Semantic Web Standards’. RDF - Semantic Web Standards. 2015. http://www.w3.org/RDF/.

Reidenberg, Joel. 2005. ‘Technology & Internet Jurisdiction’. *University of Pennsylvania Law Review*, Fordham Law Legal Studies, 153: 1951–74.

Resnick, Paul, R. Kelly Garrett, Travis Kriplean, Sean A. Munson, and Natalie Jomini Stroud. 2013. ‘Bursting Your (Filter) Bubble: Strategies for Promoting Diverse Exposure’. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*, 95. San Francisco, USA: ACM Press. https://doi.org/10.1145/2441955.2441981.

Restore NV. 2016. ‘About Us’. Restore NV. 2016. https://www.restore.eu/en/homepage.

———. 2017. ‘Restore Deploys Smart Appliances in World-Leading Smart City Project’. Restore NV. 23 January 2017. https://www.restore.eu/en/news/press-release/restore-deploys-smart-appliances-in-world-leading-smart-city-project.

Rey, P.J. 2012. ‘Alienation, Exploitation, and Social Media’. *American Behavioral Scientist* 56 (4): 399–420. https://doi.org/10.1177/0002764211429367.

Reyden, Thomas A.C. 2011. ‘Philosophy of Technology’. *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/technolo/#SH4a.

Rijt, Jan-Willem van der. 2012. *The Importance of Assent: A Theory of Coercion and Dignity*. Library of Ethics and Applied Philosophy 25. Dordrecht: Springer.

Ritchey, Tom. 2009. ‘17. Morphological Analysis’. In *Futures Research Methods*. Washington, DC.

Ritzer, George. 2011. *Sociological Theory*. 8. ed. New York: McGraw-Hill.

Robinson, Howard. 2011. ‘Dualism’. Stanford Encyclopedia of Philosophy. 2011. http://plato.stanford.edu/entries/dualism/.

Rolim, Carlos Oberdan, Anubis G. Rossetto, Valderi R.Q. Leithardt, Guilherme A. Borges, Cláudio F.R. Geyer, Tatiana F.M. dos Santos, and Adriano M. Souza. 2016. ‘Situation Awareness and Computational Intelligence in Opportunistic Networks to Support the Data Transmission of Urban Sensing Applications’. *Computer Networks* 111 (December): 55–70. https://doi.org/10.1016/j.comnet.2016.07.014.

Rubin, Anita. 2011. 'Living in the Age of Emotional Rationality: Wendell Bell, Social Media and the Challenges of Value Change'. *Futures* 43 (6): 583–589.

Rubinstein, Ira S. 2013. 'Big Data: The End of Privacy or a New Beginning?' *International Data Privacy Law* 3 (2): 74–87.

Rumbaugh, James, Michael Blaha, William Premerlani, Frederick Eddy, and William E Lorensen. 1991. *Object-Oriented Modeling and Design*. Englewood Cliffs, NJ: Prentice-Hall.

'Safe Cities'. 2014. Brochure. Cobham PLC.

Sandoval, Marisol. 2012. 'A Critical Empirical Case Study of Consumer Surveillance on Web 2.0'. In *Internet and Surveillance*, edited by Christian Fuchs, Marisol Sandoval, Boersma Kees, and Anders Albrechtslund, 147–69. Routledge Studies in Science, Technology and Society. New York, N.Y: Routledge.

Santos Brito, Kellyton dos, Frederico Araujo Durao, Vinicius Cardoso Garcia, and Silvio Romero de Lemos Meira. 2013. 'How People Care about Their Personal Data Released on Social Media'. In *Proceedings of 2013 Eleventh Annual International Conference on Privacy, Security and Trust (PST)*, 111–18. Tarragona: IEEE. https://doi.org/10.1109/PST.2013.6596044.

Schmidt, Glen M, and Cheryl T Druehl. 2008. 'When Is a Disruptive Innovation Disruptive?' *Journal of Product Innovation Management* 25 (4): 347–369.

Schmitz, Amy J. 2015. 'Secret Consumer Scores and Segmentations: Separating "Haves" from "Have-Nots"'. *Michigan State Law Review* 2014 (5): 1411.

Schneewind, Jerome B. 2007. *The Invention of Autonomy: A History of Modern Moral Philosophy*. 7th ed. Cambridge: Cambridge Univ. Press.

Schneider, G Michael, and Judith Gersting. 2010. *Invitation to Computer Science*. 5th ed. Boston, MA: Cengage Learning.

Schneier, Bruce. 2015. *Data and Goliath: The Hidden Battles to Capture Your Data and Control Your World*. New York, NY: Norton.

Scholz, Trebor, ed. 2013. *Digital Labor: The Internet as Playground and Factory*. New York: Routledge, Taylor & Francis Group.

Schwab, Klaus. 2016. *The Fourth Industrial Revolution*. Geneva, Switzerland: World Economic Forum.

Scott, Ridley. 1982. *Blade Runner*. Film. Warner Brothers.

Service Systems Group. 2015a. 'HAT Briefing Paper 1 : Engineering a Market for Personal Data : The Hub-of-All-Things (HAT)'. Working Paper. WMG Service Systems Research Group Working Paper Series. Coventry: Warwick Manufacturing Group. http://wrap.warwick.ac.uk/65605/.

———. 2015b. 'HAT Briefing Paper 2 : The Hub-of-All-Things (HAT) Economic Model of the Multi-Sided Market Platform and Ecosystem'. Working Paper. WMG Service Systems Research Group Working Paper Series. Coventry: Warwick Manufacturing Group. http://wrap.warwick.ac.uk/65607/.

Sharon, Tamar. 2017. 'Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare'. *Philosophy & Technology* 30 (1): 93–121. https://doi.org/10.1007/s13347-016-0215-5.

Sherman, Richard C. 2001. 'The Mind's Eye in Cyberspace: Online Perceptions of Self and Others'. In *Towards CyberPsychology*, edited by Giuseppe Riva and Carlo Galimberti, 73–72. Studies in New Technologies and Practices in Communication. Oxford: IOS Press.

Shifman, Limor, and Asaf Nissenbaum. 2017. 'Internet Memes as Contested Cultural Capital: The Case of 4chan's/b/Board'. *New Media & Society* 19 (4): 483–501.

Singh, Jatinder. 2013. 'Critical Appraisal Skills Programme'. *Journal of Pharmacology and Pharmacotherapeutics* 4 (1): 76–77. https://doi.org/10.4103/0976-500X.107697.

Smart, Paul, Elena Simperl, and Nigel Shadbolt. 2014. 'A Taxonomic Framework for Social Machines'. In *Social Collective Intelligence*, edited by Emilio Mordini, Vincenzo Maltese, Michael Rovatsos, Anton Nijholt, and James Stewart, 51–85. Springer.

Smith, Brendan. 2012. 'Inappropriate Prescribing'. *Monitor on Psychology* 43 (6): 36.

Smith, Merritt Roe, and Leo Marx, eds. 1994. *Does Technology Drive History?: The Dilemma of Technological Determinism*. Cambridge, Mass.: MIT Press.

Söderström, Ola, Till Paasche, and Francisco Klauser. 2014. 'Smart Cities as Corporate Storytelling'. *City* 18 (3): 307–320.

Solove, Daniel J. 2004. *The Digital Person: Technology and Privacy in the Information Age*. Ex Machina: Law, Technology, and Society. New York, NY [u.a.]: New York Univ. Press.

Soltani, Ashkan, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris Jay Hoofnagle. 2010. 'Flash Cookies and Privacy.' In *AAAI Spring Symposium: Intelligent Information Privacy Management*, 2010:158–163.

Solveforx. 2016. 'Google Self-Driving Car Project Website'. Google Self-Driving Car Project. 2016. https://www.google.com/selfdrivingcar.

Soraker, Johnny Hartz, and Philip Brey. 2007. 'Ambient Intelligence and Problems with Inferring Desires from Behaviour'. *International Review of Information Ethics* 8: 7–12.

Sparrow, Robert. 2007. 'Killer Robots'. *Journal of Applied Philosophy* 24 (1): 62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x.

Spiekermann, Sarah. 2015. *Ethical IT Innovation: A Value-Based System Design Approach*. Boca Raton, Fl.: Taylor & Francis.

Spielberg, Steven. 2002. *Minority Report*. Film. Twentieth Century Fox Film Corporation.

Stahl, Bernd Carsten. 2011a. 'IT for a Better Future: How to Integrate Ethics, Politics and Innovation'. *Journal of Information, Communication and Ethics in Society* 9 (3): 140–56. https://doi.org/10.1108/14779961111167630.

———. 2011b. 'Project Final Report'. ETICA Project.

———. 2011c. 'What Does the Future Hold? A Critical View of Emerging Information and Communication Technologies and Their Social Consequences'. In *Researching the Future in Information Systems*, 59–76. Springer. http://link.springer.com/chapter/10.1007/978-3-642-21364-9_5.

———. 2013. 'ETICA Report Summary'. European Commission : CORDIS : Report Summaries : : Report Summaries. 18 January 2013. http://cordis.europa.eu/result/report/rcn/56066_en.html.

Stahl, Bernd Carsten, Richard Heersmink, Philippe Goujon, Catherine Flick, Jeroen van den Hoven, Kutoma Wakunuma, Veikko Ikonen, and Michael Rader. 2010. 'Identifying the Ethics of Emerging Information and Communication Technologies: An Essay on Issues, Concepts and Method'. *International Journal of Technoethics (IJT)* 1 (4): 20–38.

Stahl, Bernd Carsten, B. Mittelstat, and N. B. Fairweather. 2013. 'PHM-Ethics and ETICA: Complementary Approaches to Ethical Assessment'. In *Interdisciplinary Assessment of Personal Health Monitoring*, edited by S Schmidt and O Rienhoff, 117–36. Studies in Health and Technology Informatics. Amsterdam: IOS Press.

State Council of People's Republic of China. 2014. 'Planning Outline for the Construction of a Social Credit System (2014-2020)'. Planning Outline GF No. (2014)21. China: State Council of People's Republic of China. https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/.

Statistica. 2016. 'Statistics and Facts about Tesla'. The Statistics Portal. 2016. http://www.statista.com/topics/2086/tesla/.

Stevens, W. Richard, and Gary R. Wright. 1994. *TCP/IP Illustrated Volume 1: The Protocols*. 2nd ed. Vol. 1. 2 vols. Addison-Wesley Professional Computing Series. Reading, Mass.: Addison-Wesley.

Stige, Brynjulf, Kirsti Malterud, and Torjus Midtgarden. 2009. 'Toward an Agenda for Evaluation of Qualitative Research'. *Qualitative Health Research* 19 (10): 1504 – 1516. https://doi.org/10.1177/1049732309348501.

Stole, Inger L. 2014. 'Persistent Pursuit of Personal Information: A Historical Perspective on Digital Advertising Strategies'. *Critical Studies in Media Communication* 31 (2): 129–33. https://doi.org/10.1080/15295036.2014.921319.

Stoljar, Natalie. 2015. 'Feminist Perspectives on Autonomy'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2015. http://plato.stanford.edu/archives/fall2015/entries/feminism-autonomy/.

Stoljar, Natalie, and Catriona MacKenzie, eds. 2000. *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self.* New York: Oxford University Press.

Strawson, Galen. 1994. 'The Impossibility of Moral Responsibility'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 75 (1/2): 5–24.

Suler, John. 2004. 'The Online Disinhibition Effect'. *CyberPsychology & Behavior* 7 (3): 321–26. https://doi.org/10.1089/1094931041291295.

Sweeney, Latanya. 2013. 'Discrimination in Online Ad Delivery'. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2208240.

Tabak, Edin. 2015. *Information Cosmopolitics: An Actor-Network Theory Approach to Information Practices*. Chandos Information Professional Series. Waltham, MA: Chandos Publishing.

Tavani, Herman T. 2007. 'PHILOSOPHICAL THEORIES OF PRIVACY: IMPLICATIONS FOR AN ADEQUATE ONLINE PRIVACY POLICY'. *Metaphilosophy* 38 (1): 1–22. https://doi.org/10.1111/j.1467-9973.2006.00474.x.

Taylor, Nigel. 2005. *Urban Planning Theory Since 1945*. London: SAGE Publications.

Tene, Omar, and Jules Polonetsky. 2012. 'To Track or "Do Not Track": Advancing Transparency and Individual Control in Online Behavioral Advertising'. *Minnesota Journal of Law, Science & Technology* 13 (1): 281–358.

Tene, Omer, and Jules Polonetsky. 2013. 'Big Data for All: Privacy and User Control in the Age of Analytics'. *Northwestern Journal of Technology and Intellectual Property* 11 (5): 239–72.

The Select Committee on Digital Skills. 2015. 'Make or Break: The Digital Future'. Report of Session 2014–15. UK: The Authority of the House of Lords. http://www.publications.parliament.uk/pa/ld201415/ldselect/lddigital/111/111.pdf.

Tibbits, Skylar, and Kenny Cheung. 2012. 'Programmable Materials for Architectural Assembly and Automation'. *Assembly Automation* 32 (3): 216–225.

Tomlinson, John. 1991. *Cultural Imperialism: A Critical Introduction*. Baltimore, Md.: Johns Hopkins University Press.

Tötterman, Henrik, and Jan Sten. 2005. 'Start-Ups: Business Incubation and Social Capital'. *International Small Business Journal* 23 (5): 487–511. https://doi.org/10.1177/0266242605055909.

Trist, Eric L. 1981. *On Socio-Technical Systems*. Ontario: Ontario Ministry of Labour.

Turkle, Sherry. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books. http://public.eblib.com/choice/publicfullrecord.aspx?p=684281.

Turow, Joseph. 2011. *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your World*. New Haven: Yale University Press.

Ugolev, Boris N. 2014. 'Wood as a Natural Smart Material'. *Wood Science and Technology* 48 (3): 553–568.

UNESCO Secretariat. 2014. 'Internet Universality'. United Nations Educational, Scientific and Cultural Organization (UNESCO). http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/news/internet_universality_en.pdf.

Ustaran, Eduardo, and Mac Macmillan. 2015. 'Future-Proofing Privacy'. London: Hogan Lovells.

Vaidhyanathan, Siva. 2012. *The Googlization of Everything:(And Why We Should Worry)*. Berkeley, CA.: University of California Press.

Van Eck, Nees Jan, and Ludo Waltman. 2007. 'VOS: A New Method for Visualizing Similarities Between Objects'. In *Advances in Data Analysis: Proceedings of the 30th Annual Conference of the German Classification Society*, 299–306. Springer. http://link.springer.com/chapter/10.1007%2F978-3-540-70981-7_34.

———. 2009. 'Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping'. *Scientometrics*. http://link.springer.com/article/10.1007/s11192-009-0146-3/fulltext.html.

Varela, Francisco G, Humberto R Maturana, and Ricardo Uribe. 1981. 'Autopoiesis: The Organization of Living Systems, Its Characterization and a Model'. *Cybernetics Forum* X (2–3): 7–13.

Veikko, Ikonen, Minni Kanerva, and Panu Kouri. 2009. 'D.1.1 Heuristics & Methodology Report - Final'. D.1.1. ETICA Project.

Veikko, Panu Kouri, and Ikonen. 2011. 'D.1.3 Emerging Technologies Workshop Report'. D.1.3. ETICA Project.

Verghese, Preeti. 2001. 'Visual Search and Attention: A Signal Detection Theory Approach'. *Neuron* 31 (4): 523–535.

Verizon. 2014. '2014 Data Breach Investigations Report'. Annual Verizon Data Breach Investigations Report. Verizon Enterprise Solutions.

Verwijmeren, Thijs, Johan C Karremans, Wolfgang Stroebe, and Daniël HJ Wigboldus. 2011. 'The Workings and Limits of Subliminal Advertising: The Role of Habits'. *Journal of Consumer Psychology* 21 (2): 206–213.

Vinod Kumar, T. M., ed. 2015. *E-Governance for Smart Cities*. Advances in 21st Century Human Settlements. Singapore: Springer Singapore. http://link.springer.com/10.1007/978-981-287-287-6.

Von Bertalanffy, Ludwig. 1968. *General System Theory*. Vol. 41973. New York, N.Y: George Braziller, Inc.

Voorsluys, William, James Broberg, and Rajkumar Buyya. 2011. 'Introduction to Cloud Computing'. In *Cloud Computing: Principles and Paradigms*, 1–44. New York: Wiley Press.

Wang, Shaojung Sharon. 2013. '"I Share, Therefore I Am": Personality Traits, Life Satisfaction, and Facebook Check-Ins'. *Cyberpsychology, Behavior, and Social Networking* 16 (12): 870–77. https://doi.org/10.1089/cyber.2012.0395.

Wayne, Logan Danielle. 2012. 'The Data-Broker Threat'. *Journal of Criminal Law and Criminology* 102 (1): 253–82.

Westin, Alan F. 1970. *Privacy and Freedom*. London: Bodley Head.

Whitman, James Q. 2004. 'The Two Western Cultures of Privacy: Dignity versus Liberty'. *The Yale Law Journal* 113 (6): 1151. https://doi.org/10.2307/4135723.

Wilske, S., and T. Schiller. 1997. 'International Jurisdiction in Cyberspace: Which States May Regulate the Internet?' *Federal Communications Law Journal* 50: 119–78.

Wilson, Edward O. 1978. *On Human Nature*. Cambridge, Mass.: Harvard University Press.

Wilson, Mark I, and Kenneth E Corey. 2011. 'Approaching Ubiquity: Global Trends and Issues in ICT Access and Use'. *Journal of Urban Technology* 18 (1): 7–20.

Winfield, Alan FT, Christian Blum, and Wenguo Liu. 2014. 'Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection'. In *Advances in Autonomous Robotics Systems*, 85–96. Springer.

Wood, Allen W. 2006. *Karl Marx*. 2. ed. Arguments of the Philosophers. New York: Routledge.

XML Working Group. 2015. 'XML Core Working Group Public Page'. XML Core Working Group Public Page. 2015. http://www.w3.org/XML/Core/#Publications.

Yao, Mike Z., and Daniel G. Linz. 2008. 'Predicting Self-Protections of Online Privacy'. *CyberPsychology & Behavior* 11 (5): 615–17. https://doi.org/10.1089/cpb.2007.0208.

York, Dan. 2016. *Overview of the Digital Object Architecture (DOA)*. Geneva, Switzerland: ISOC.

Zekos, G. I. 2006. 'State Cyberspace Jurisdiction and Personal Cyberspace Jurisdiction'. *International Journal of Law and Information Technology* 15 (1): 1–37. https://doi.org/10.1093/ijlit/eai029.

Zhang, Jiajie, and Vimla L Patel. 2006. 'Distributed Cognition, Representation, and Affordance'. *Pragmatics & Cognition* 14 (2): 333–34.

Zhao, Shanyang, Sherri Grasmuck, and Jason Martin. 2008. 'Identity Construction on Facebook: Digital Empowerment in Anchored Relationships'. *Computers in Human Behavior* 24 (5): 1816–36. https://doi.org/10.1016/j.chb.2008.02.012.

Zimbardo, Philip G, and Michael R Leippe. 1991. *The Psychology of Attitude Change and Social Influence*. Philadelphia: Temple University Press.

Zygiaris, Sotiris. 2013. 'Smart City Reference Model: Assisting Planners to Conceptualize the Building of Smart City Innovation Ecosystems'. *Journal of the Knowledge Economy* 4 (2): 217–231.

13. Appendix of Published Works

The following pages contain the published papers as printed in the journals. As the book chapter, "Binding the Smart City Human-Digital System with Communicative Processes" is still in press, no published version exists, so it is not included in this section.

Key Dialectics in Cloud Services

Brandt Dainow
Department of Computer Science, Maynooth University
Kildare, Co. Kildare
Ireland
+353 86 248 2846
brandt.dainow@nuim.ie

ABSTRACT

This paper will identify three central dialectics within cloud services. These constitute defining positions regarding the nature of cloud services in terms of privacy, ethical responsibility, technical architecture and economics. These constitute the main frameworks within which ethical discussions of cloud services occur.

The first dialectic concerns the question of whether it is it essential that personal privacy be reduced in order to deliver personalised cloud services. I shall evaluate the main arguments in favour of the view that it is. To contrast this, I shall review Langheinrich's *Principles of Privacy-Aware Ubiquitous Systems* [24]. This offers a design strategy which maintains functionality while embedding privacy protection into the architecture and operation of cloud services.

The second dialectic is concerned with the degree to which people who design or operate cloud services are ethically responsible for the consequences of the actions of those systems, sometimes known as the "responsibility gap." I shall briefly review two papers which argue that no one is ethically responsible for such software, then contrast them with two papers which make strong arguments for responsibility. I shall show how claims for no responsibility rest on very narrow definitions of responsibility combined with questionable conceptions of technology itself.

The current shape of cloud services is dominated by a tension between open and closed systems. I shall show how this is reflected in architecture, standards and organisational models. I will then examine alternatives to the current state of affairs, including recent developments in support of alternative business models at government level, such as the House of Lords call for the Internet to be treated as a public utility (The Select Committee on Digital Skills, 2015).

CATEGORY

K.4.1 **[Public Policy Issues]**: Ethics

GENERAL TERMS

Design. Human Factors.

KEYWORDS

Cloud services, ethics, privacy, security, privacy by design, personalization, filter bubble

INTRODUCTION

This paper explores dialectics within debates regarding key ethical issues pertaining to cloud services. These issues concern privacy, responsibility for the actions of systems and the development of monopoly service providers. Between them these concerns largely dictate the shape and capabilities of current and future cloud-based services. I shall show how the current state of affairs is dominated by a sense of lack of agency in terms of doing things differently from the current reflexive practice, an assumption that no alternatives to current practice are possible. This paper will attempt to organise the key concerns with cloud services by arranging them into three dialectical axes:

- The nature of the relationship between personal privacy and service provision.
- The degree to which people who build or operate cloud-based services are ethically responsible for the actions or effects of those services.
- The nature of the marketplace for those services.

Since this paper considers cloud services in the broadest sense, it is appropriate to commence with the definition of cloud computing used in this analysis. The *US National Institute of Standards and Technology Special Publication 800-145* defines the essential characteristics of cloud computing as being:

- The ability to provide services whenever desired without human intervention.
- Being available to a wide range of client devices via networking technology.
- The "virtualisation" of computing resources, such that digital operations are not linked to specific servers or locations.
- Scalability – the capability of the systems to scale up or down in response to changes in demand (a necessary corollary of virtualisation).
- Often, but not necessarily, Software as a Service. [29]

Clearly this definition applies to many, if not most, internet systems and digital services, not merely to the virtualisation of server functions previously found in the traditional client-server network. Under this view, Facebook and Google search are both cloud services. I think this is both valid and important - confining discussion of cloud computing to data processing or file storage functions limits discussion to a few contingent uses of a wider system and obscures the essential factors we need to consider.

PRIVACY VERSUS SECURITY

Our first axis is the necessity versus the contingency of reductions to privacy under new digital services. That is to say, there is one body of opinion which holds that the erosion of personal privacy is a necessary and unavoidable consequence of, or precondition for, the delivery of digital services. These positions tend to be a reflexive response within the development community, rarely stated formally, and is a minority view in the literature, as a result of which detailed arguments as to why privacy *must* be reduced to enable cloud services are scarce. However, Lucas Bergkamp's paper, *The Privacy Fallacy* [5] marshals all the arguments in this camp.

Bergkamp argues there should be no privacy protection of any form because preservation of personal privacy is harmful to society in many ways. He provides five main arguments; there is no need for data privacy, data protection reduces individual freedom, personal privacy is contrary to economic growth, EU data legislation is unenforceable and the EU's data protection regimes put it out of step with the rest of the planet. I will now explore each of these in more depth:

Bergkamp argues there is no need for data protection or digital privacy because no one wants it and it serves no purpose. He states there is no evidence anyone has ever been harmed by privacy violations or personalization of services based on personal data. He does not provide any evidence for this and it is contradictory to the reported activity of many data protection authorities. For example, in 2014 the Data Commissioner of Ireland received 2,264 data breach notifications, investigated 960 complaints and launched 162 prosecutions. Half (53%) of complaints involved disclosing personal data inappropriately, such as disclosure of personal financial data to relatives or the listing of email addresses and passwords on public websites [34]. The *Verizon 2014 Data Breach Investigations Report* [46] covers 63,000 data violations across 93 countries in 2014. It highlights financial theft and the cost of dealing with a breach, such as cancelling credit cards, as the main harms to the individual. Other research exists to show harm from less obvious privacy violations. *RT@Iwantprivacy: Widespread Violation of Privacy Settings in the Twitter Social Network* details harm from privacy violations in Twitter when people reuse private tweets in public [28]. *Privacy Violations Using Microtargeted Ads* [21] details harm from privacy violations in Facebook. Privacy violations has also been shown to harm the companies themselves. *How Privacy Flaws Affect Consumer Perception* [2] shows how privacy breaches reduce the chance people will buy from a company, while *Is There a Cost to Privacy Breaches?* [1] shows how privacy violations reduce a company's share price. Studies also exist to show harm from personalization of advertising and news. The research findings of Sweeney's *Discrimination in Online Ad Delivery* [42] reveal how racial stereotyping in ad personalization harms Afro-Americans in many ways, including job prospects and access to financial services. *Bursting Your Filter Bubble* [38] shows harm from news personalization, while the famous Facebook news manipulation study, *Experimental Evidence of Massive-scale Emotional Contagion through Social Networks* [22] shows how personalizing news feeds to contain more negative contents can depress people. Bergkamp's proposition that there has never been any harm from privacy violations or personalization appears to be contradicted by such evidence.

Bergkamp also argues there is no need for data protection because no one wants it. He argues that people don't realise that data protection prevents personalization, but that when they do, they always prefer personalization over data protection. He does not cite any evidence for this. By contrast, Culnan's 1993 study of personalization in shopping, *How Did They Get My Name?* [14] shows that when offered the choice, the people he surveyed preferred privacy over personalisation. More recently, the *2013 Comres Big Brother Watch Survey* [11] polled 10,000 people in nine EU countries to find 75% were concerned about privacy and wanted data protection regulations, while 45% believed they were being harmed by corporate data practices.

Bergkamp also argues there is no need to regulate sale of personal data because companies never sell it. However, there is, in fact, a huge industry in the sale and aggregation of personal data, as the 2014 Federal Trade Commission's investigation into data brokers found [7,17].

Bergkamp argues that personalization results in cheaper prices. However, he does not cite any empirical evidence for this or reasons why it should be so. He cites as evidence a statement made by Fred Cate, Professor of Law at Indiana University, that personalization results in cheaper prices, but this was a statement made to a Congressional committee, not a research finding. Prof. Cate's own list of publications does not include any research into personalization, his speciality is data protection law. Later in the paper Bergkamp states that data protection costs money, and that it is so burdensome and expensive that businesses can only survive by ignoring their legal obligations. One may surmise he believes this is the cause of higher prices to consumers, though he does not explictly say so. However, research like Sweeney's *Discrimination in Online Ad Delivery* [42] shows how personalization actually increases costs to Afro-American consumers in the USA, while Turow's *The Daily You: How the New Advertising Industry is Defining Your Identity and Your World* [44] shows how personalization can reduce or increase prices, depending on whether you are the consumer companies want or not.

Bergkamp also argues that privacy protection increases identity theft because data protection makes it harder to tell if someone really is who they claim to be. He does not cite any evidence for this and it seems counter-intuitive. Given that privacy protection reduces access to the personal data necessary for identity theft, such protection could be presumed to make it harder to commit, so one could argue the exact opposite of Bergkamp in the absence of any research. Bergkamp's position here allies with his arguments elsewhere in his paper that we all need to know as much as possible about each other in order to protect ourselves from one another, and that privacy directly prevents this. He states that one problem with privacy protection is that it allows an individual to control what they disclose to the world. He does not explicitly say this is a bad thing, but it is clearly implied from his usage. Here it is worth noting research showing the reverse, that lack of privacy restricts human freedom. For example, knowledge one is being watched on the internet has been shown to have a chilling effect on what people say [4] and what they search for [25], even when engaging in legal and socially acceptable activity.

Bergkamp claims there is a vast amount of money to be made acquiring and selling personal data, despite his earlier claim that there are no businesses selling it. He provides no evidence for this economic activity, but the claim is supported elsewhere. For example, in 2013 the OECD estimated the personal data of each Facebook user to range from $US40/year to $US400 [33]. Bergkamp claims that this data market alone is sufficient reason to remove privacy protections. However, the mere presence of economic activity does not, in and of itself, mean we should encourage it. There is a vast amount of money to be made in drug smuggling, but no one uses that as an argument for encouraging it.

Bergkamp also states that the EU's data legislation is unenforceable. He says the very concept of personal privacy is too vague to support regulation and that the regulations cannot properly specify what constitutes personal data. Furthermore, he says, each privacy incident must be judged on its own merits. He does not explain how judging a case on its own merits is a problem. Each and every infraction of the law is judged individually, so arguing that this is also the case for privacy issues does not, in and of itself, constitute a sign of poor legislation. Furthermore, it is difficult to imagine what the alternative would be if a regulator or judge was not allowed to consider the specific details of each case they were trying to adjudicate.

As stated earlier, Bergkamp believes that data protection is so onerous that no business can do it properly and survive financially. He claims the only outcome is that data regulations are never enforced. Clearly the many cases of prosecution for privacy violations are not accounted for in this argument. Bergkamp also states that EU data protection legislation is founded on a misunderstanding of how business works, but does not provide any further details regarding the nature of the misunderstanding or what the reality truly is.

Bergkamp's paper also states privacy protection damages society because it involves the government paternalistically interfering in people's relations with each other in a misguided attempt to stop people hurting each other. Such an argument can also be said of laws against violence and theft, so the logical consequence of such a position is that we should move to a state of complete anarchy. However, Bergkamp does not address this implication. Instead he goes on to state that government's should never restrict any information under any circumstances. Again no reasons are provided to justify this proposition. Such a broad statement can also be used as an argument in favour of making child pornography freely available, so some additional clarification would seem appropriate.

Finally, Bergkamp claims that EU data protection legislation is out of step with the rest of the world. He does not provide any evidence to support this, but he clearly thinks this is a bad thing and grounds for abandoning data protection. This is a questionable claim. The EU's data protection regime was intentionally built to accord with pre-existing OECD guidelines, which were first developed in 1980 [32].

Bergkamp never states that it is technically impossible to maintain privacy while extending cloud services. His arguments are merely that we should not. My position is that there is no necessary and unavoidable relationship between privacy consequences and functionality. One does not *have* to reduce privacy in order to extend services. Rather, it is always a question of choice, either in how the system is constructed or in the type of business model under which it operates, and there are *always* alternatives. It may be that some of those alternatives are more expensive than the privacy-reducing models, or that alternatives are more technically challenging. However, that, in and of itself, is not an argument for the necessity of privacy-reducing models, but rather an argument underpinning a particular business model or software approach.

Currently those who are building cloud-based services most commonly work on the basis that privacy is exchanged for digital services. However, there is also a growing body of those seeking to develop alternatives, in terms of governance or business model or in terms of code. The most notable is the Privacy by Design movement. However, most of the Privacy by Design material is so vague as to be little more than statements of intent. For example, IBM claim to have moved to Privacy by Design by doing nothing more than implementing awareness training and building an internal system for reporting data breaches [35]. Here Langheinrich's paper, *Principles of Privacy-Aware Ubiquitous Systems* [24] stands out as the exception, being a concrete statement of specific technical design principles which genuinely do embed privacy considerations into the technical architecture. Langheinrich's paper shows it is possible to build robust systems which have privacy protection embedded within the design and operation of the system.

It is notable that Langheinrich has practical experience in the design of privacy systems, being one of the authors of the W3C's technical standard, *Platform for Privacy Preferences*, or PPP [13]. The PPP standard enables browsers to hold the user's preferences for what data they will allow a website to gather. The server component of PPP allows the web server to list its own data-gathering practices. PPP then enables the browser to compare the web server's practices with the user's preferences. The system provides for warnings to the user and for compact and rapid communication between client and server of data practices. The system was supported in Microsoft Internet Explorer 6 when it first emerged, but lack of support by website owners means PPP is largely unused today.

The first of Langheinrich's principles is the Principle of Openness, or "Notice." This simply states that no device or service should gather data about someone without telling them. Here he makes reference to PPP as providing a digital vocabulary which could be used to programmatically describe what data is being gathered, for what purpose and by whom. This is paired with the second principle, the Principle of Consent, which encodes the legal necessity for informed consent. A system must allow for someone to opt out of being tracked or recorded, and do so without denying service on a "take it or leave it" basis. Thus, for example, buildings would need to disable tracking for some people and not simply refuse them entry.

The third principle is termed "Anonymity and Pseudonymity." This states that people must have the option to remain anonymous. The issue here is that some services are only possible if they know a user's identity and history. Here Langheinrich introduces pseudonymity. Under this system a person may have a unique identifier of some form, such as a cookie or RFID chip, which anchors the data systems and forms the index key to their personal data history. However, this identifier contains no personally identifiable information and is discardable at any time. Furthermore, such a system permits people to have multiple pseudonymous ID's and so prevent aggregation of disparate activities by data brokers. It is noteworthy that EU data regulations have recently been updated to add the category of pseudonymous identity between personal and anonymous data [48].

Langheinrich's fourth principle of "Proximity and Locality" limits the scope of data collection. Looking to a future in which people have many devices capable of recording their surroundings, the principle of proximity states that these devices can only operate in the proximity of their owner. This prevents people leaving devices to record data unseen, then returning for them later. Of wider application is the principle of locality; devices should not transmit data any further than absolutely necessary to fulfil their functions. For example, Samsung's voice-activated TV's transmit all conversations they hear to Samsung's central servers. Voice commands are interpreted there and the appropriate command then sent back to the TV. All conversation recordings are stored permanently for later analysis [49]. Under Langheinrich's principles the TV would have been designed so that it did not need to involve cloud services. Voice recognition chips have been around for 20 years and could have be used instead.

The fifth principle is the "Need for Security," in which Langheinrich advocates various levels of security depending on the nature of the data. More importantly, he illustrates how the previous principles themselves enhance security. If data is not being transmitted many security problems simply vanish. Similarly, if data is not linked to an identifiable individual, but only to a pseudonymous ID, unauthorised access has less potential for harm.

Langheinrich's final principles are the principles of "Collection and Use Limitation." These state that data collectors should only collect data for a specific purpose and not store it, as Samsung TV does, in case they want to use it in the future. Secondly, they should only collect the data they need in order to fulfil their task and nothing more. Finally, they should only keep data as long as it is necessary for the purpose. While these appear primarily legislative principles, they can be embodied in technical design through the use of the earlier principles. For example, if data is housed in the user's devices in accordance with the principle of locality, then the user can impose usage and storage limitations themselves.

Langheinrich's principles, if implemented, would solve many privacy concerns, enhance security and actually make many applications of ubiquitous and cloud services easier to construct. What they show is that it is perfectly possible to design cloud services in a manner which enhances both security and privacy at the same time, while permitting all the personalization necessary. They place control of personal data firmly in the hands of the user without compromising technical operations in any way. In fact, their reduced dependence on permanent access to centralised services makes them more robust and reduces the burden of traffic on the internet. These principles are easy to understand and yet produce powerful architectures. They offer a practical and detailed response to the reflexive position that personalized cloud services must reduce privacy. In doing so they provide concrete evidence that it would be possible to move cloud service evolution into a path which fulfils all its potential, yet enhances privacy and security at the same time. Langheinrich's design principles demonstrate that the reduction of privacy in cloud services is a choice, not a necessity.

ETHICAL RESPONSIBILITY

Our second axis is concerned with the degree to which people who design, build or operate cloud services are ethically responsible for the consequences of the actions of those systems. This question does not arise with regard to all cloud services, but only with the rising generation of autonomous services which process personal data in order to deliver personalised services, such as personalised search results, product recommendations and news feeds. In the near future we will see the rise of more intelligent and more life-critical personalised services, most notably with bio-implantation and other medical services [19]. The question is primarily one of who is responsible when such autonomous services make decisions which result in harm, but where these decisions are not the result of faulty design or incorrect data.

The competing positions are that, on the one hand, programmers and operators are not ethically responsible for the actions of autonomous systems, versus a view that they are. It is difficult to argue that the person holding a hammer is not ethically responsible for the consequences of whatever happens when the hammer hits something because the hammer is totally under the control of the user. However, with large industrially-produced complex automated systems, especially those that include some form of AI functionality, arguments emerge in favour of the position that those who build the systems are not ethically responsible for the decisions those systems make. This argument will no doubt be exacerbated the more powerful and the more intelligent and autonomous these systems become. This issue is discussed most frequently with regard to autonomous military systems, whose lethality makes the question of ethical responsibility both stark and urgent. However, the question is just as pertinent for any form of autonomous system, including those cloud-based personalization systems already in operation.

Andreas Matthias' paper, *The Responsibility Gap* [26] offers a fairly straightforward account based on philosophical logic to support the position that programmers are not responsible for the actions of their autonomous systems, while Robert Sparrow's *Killer Robots* [39] presents the same conclusion via an examination of the practicalities of creating and deploying autonomous systems. Both take the position that no one at all is ethically responsible for the actions of autonomous agents.

Sparrow's argument is based on his particular understandings of the terms 'autonomy' and 'responsibility.' My view is that he defines these terms in such a way as to make any contrary conclusion impossible. Early in the paper, he defines autonomy as being free from external causation:

> "Where an agent acts autonomously, then, it is not possible to hold anyone else responsible for its actions. In so far as the agent's actions were its own and stemmed from its own ends, others cannot be held responsible for them. Conversely, if we hold anyone else responsible for the actions of an agent, we must hold that, in relation to those acts at least, they were not autonomous." [39:65–66]

Sparrow does not defend this definition of autonomy. However, once it has been defined this way, it becomes a matter of logical necessity that there is no ethical responsibility by the programmers or controllers. It is also worth noting that Sparrow uses autonomy in an absolute sense, as if the agent were free from all influence except their prior experience. In particular, he does not recognise the environment, the capabilities of the device or its internal structures as having any impact on decision-making. He argues the programmer cannot be responsible because the essence of an autonomous system is that it will make unpredictable decisions. He argues the controller of the system is not responsible because they could not anticipate what it would do any better than the programmer. In both cases, he ignores the fact the system is designed to perform a particular role in a particular environment. A software agent is not free to do just anything, it can only recognise inputs of a type it has been designed for, and has a relatively limited range of actions it can take, and can only operate in a specific type of environment. A share-dealing system cannot walk the dog or assess your exercise regime. The type of decisions an autonomous system may make, and the range of options available to it, are not only predictable, they are the basis upon which it was designed and built - they define it. An autonomous system may make its own decisions, even alter its own programming, but its range of actions and the forms of harm it may commit are knowable in advance in virtue of the type of system it is.

Sparrow does not mention Strawson in his paper, but his conception of moral responsibility has close parallels to Strawson's influential work. Strawson's position is that no one is morally responsible for anything because no one is free from external influence [41], though the details of why are beyond the scope of this paper. Though Sparrow does not say so explicitly, his use of responsibility is clearly that one can only be responsible for specific actions. In Sparrow's view, the design of the system and the decision to use it do not carry any ethical responsibility because neither gives one the ability to predict the specifics of an individual act the system may take.

Sparrow also argues it is not possible to hold the system itself responsible because responsibility necessarily requires punishability which requires suffering. Under his definitions, something can only be morally responsible if it can be punished and something can only be punished if it can suffer. Since software systems cannot be made to suffer, they cannot be punished and so cannot be held responsible for their actions. Note here that we have switched from talk of "being responsible" to talk of "being held responsible." Here we see that Sparrow has conflated the moral state of being responsible with the social status of being eligible for punishment.

Matthias's *The Responsibility Gap* [26] also argues that no one is responsible for the decisions of autonomous software systems. His position also links responsibility to individual acts, holding that one can only be responsible if one can know the internal state of the system *and* has control of each act it takes, at least to the degree where one could prevent it. Under this analysis a programmer has no responsibility for the actions of a system once the owner takes control. The owner is not responsible because they cannot know the internal state of the system. Matthias spends some time examining different types of AI learning, showing how each makes their internal state unknowable in different ways, but the differences do not affect his final conclusion.

The narrow understanding of responsibility seen in Matthias and Sparrow is the foundation on which their arguments rest. In contrast, Miller's *Collective Responsibility and Information and Communication Technology* [30] confronts this issue by arguing there are different types of responsibility. In addition to the responsibility for individual acts which Matthias and Sparrow focus on, Miller points out we also recognise one can have "structural" responsibility by creating the conditions which made the act possible or by ordering others to take actions which eventually led to the act. Under Miller's analysis both programmers and controllers of autonomous systems take structural responsibility for every act taken by these systems. Miller then goes deeper, investigating the concept of collective responsibility. He argues that to the degree that individuals contribute something to the shape and operation of an autonomous system, so they share in responsibility for its actions. Here he acknowledges the existence of corporate responsibility, but argues that it does not provide a moral shield for the individual workers, whose individual contributions to a system's operation convey a share in collective responsibility for its actions.

Miller's approach is a step towards recognition that software agents exist within the wider context of human activity. This broader perspective is fully achieved in *Software Agents, Anticipatory Ethics, and Accountability* by Johnson et. al. [20]. Reiterating the perspective that technology is socially situated, this paper argues that the concept of any digital service as an autonomous agent is merely metaphorical; that no such system can be autonomous in the sense we apply autonomy to humans in moral debates. As such, the use of the metaphor is justifiable only by its utility. Johnson et. al. criticise the concept of any software as an autonomous agent on the grounds it generates just these ethical problems. Instead, Johnson et. al. argue we should recognise autonomous systems as elements within a larger socio-technical system, made by people and used by people for human purposes. Under this view autonomous systems are not independent entities hermetically sealed from their environments, but systems which can only be understood by reference to the context of their use. Johnson et. al. make implied use of the different forms of responsibility seen in Miller, but do not elucidate the differences. Instead they focus on the arbitrariness of delimiting technical artefacts. They deny that autonomous software agents are different in kind from any other form of automated or semi-automated device, being merely more complicated. Autonomous systems are thus merely, like hammers, extensions of human will and intent. Under this arrangement, ethical responsibility for their actions is not in any way changed by the mere fact of their complexity.

OPEN VERSUS CLOSED

Our final dialectic concerns the form and marketplace of cloud services. Here the dominating dialectic is that of open versus closed systems and open versus closed organisational contexts for such systems.

The scene for this debate is best set Eben Moglen in his presentation, *Freedom in the Cloud*, delivered to the Internet Society in 2010 [31]. Moglen argues that the internet was originally designed as a non-hierarchical peer-to-peer network. However, under the influence of the architectural model of client-server networking, the services which evolved used a smart-server-dumb-client model, in which both algorithms and data were centralised. Moglen maintains that cloud architecture works on this thin client - fat server model and does not represent a new computing architecture, merely the virtualisation of some server operations within this traditional model. These servers maintain activity logs. These logs can be mined for behavioural data. Marketing companies learned they could mine these logs to understand, predict and influence user behaviour in order to sell advertising. Moglen contends that as the perceived value of this information grew, it spurred the development of a secondary internet infrastructure of tracking services designed to add to the growing database of what we now call "user profiles."

Thus Moglen describes how an architecture which concentrates processing power and data at centralised locations promotes a concentration of both technical proficiency and economic power, while also promoting a top-down hierarchical organisational model and, in a global internet, the development of a limited number of very large monopoly service providers. This has, he argues, produced an extreme power dichotomy between those who own the services and those who use them. The business model which has come to dominate the internet is that of delivering services in exchange for spying on the users all the time. Moglen describes this state of affairs as undesirable for two reasons. Firstly, the price is too high and the services are not worth the loss of privacy. Secondly, the lack of alternative models for access to the same services makes this unfair arrangement unavoidable. He argues that we need an alternative architecture in which the data about us stored on centralised servers is instead housed in devices we own and carry with us. We can then control who accesses this data and how. He argues that this is possible with current technology.

There is an additional element of concern within Moglen's model which he hints at but does not explore. The combination of architecture and business model he describes has produced "walled gardens." These are silos of private technology and proprietary data formats which are not compatible with, or accessible by, other systems or organisations. The patent system combines with a capitalist marketplace to financially reward such behaviour. If I am the sole owner of a system everyone wants to use, I can make money. If I create a system which I give away, I do not benefit. What I therefore need to do is lock everyone into my technology, and then I will "lock in" the market [12].

The effect of this is to lock data and services into a single monopoly provider. The provider becomes the gatekeeper over the knowledge of what they do and how they do it. Users cannot migrate to a competitor without significant effort and loss. For example, if you close your account with Amazon, they will remove all the books from your Kindle [50]. You cannot therefore switch to an alternative, such as Adobe Digital Editions, without re-purchasing your entire digital library. Different legal regimes permit different levels of access inside these walled gardens, but in no case does a society have full knowledge or any substantive control. Such a system has no interest in open standards, interoperability, or a free flow of information. This lack of interoperability and open standards was why the internet and HTML were not developed by commercial enterprises. Early pre-cursors of the web tried the same walled garden approach, including America Online, CompuServe and Lotus Notes. It was only when Tim Berners-Lee gave HTML away that we broke free of this limiting system and gained the web. Berners-Lee gave it away because he saw things in exactly this way and believed that if he patented or sold HTML, it would become just another walled garden [6].

However, as companies have developed services which sit atop these communally-owned standards, so they have developed further proprietary systems. The final result is that companies have built a new layer of walled gardens and data silos on top of the open platform which is the internet [16]. The scale of the internet user base combines with a shared service delivery infrastructure to enable the rise of extremely large global monopolies, such Google, Amazon and Facebook. The result is that cloud services are portioned out amongst a limited number of very large hierarchical organisations, each of which hides its use of data from public scrutiny and uses its monopoly position and ownership of personal data as a competitive advantage [8,15,27]. The net effect is that people are locked to service providers like serfs to their lord. However, unlike in the Middle Ages, there is no competing lord to flee to if you are unhappy with your lot. This power is a concern to many. Some argue, for example, that Amazon's potential to control what books are available makes it a political institution as well as an economic one [10], while Google has consistently ranked as one of the biggest spenders on political lobbying in Washington, D.C. since 2012 [18].

Opposing this state of affairs are a disparate range of alternatives, such as Moglen and his concept of a personal server. Each alternative tends to focus on one aspect of this system, such as technical architecture or business model. Technically, the existence of the internet is based on open standards, such as TCP and IP [40], so alternatives have always been available on a technical level. Here we have the open source activists, such as the Free Software Foundation and the IETF. In addition, we have less obvious alternative architectures based on peer-to-peer (as opposed to client-server) models, such as the BOINC platform for community computing [3] and the BitTorrent protocol [36]. Standards like XML [47] and RDF [37] provide a means of breaking open walled gardens through data exchange, while people such as Chris Marsden in the UK or Robert McChesney in the USA have developed the rationale for breaking down these proprietary data silos.

McChesney argues that the development of monopolies and cartels has so dominated the internet that there has been little economic benefit for the rest of society. He argues that there is so little competition at the point of delivery that service providers constitute a cartel which should be forced into competition with not-for-profit public alternatives. He calls for the monopolistic corporations dominating important services, like Facebook and Google, to be broken into smaller competing units and subject to much more stringent and detailed state control. McChesney's argument is that the size of these corporations is so great they pose a threat to democracy itself through their power to lobby politicians, dominate online debate and skew economic development [27].

Concern over monopoly domination is addressed in a different manner by Brown and Marsden in a number of publications. Instead of seeking a solution by changing the economic structure, they focus on the proprietary data structures which form the foundation of such domination. In addition to rights such as the right to have one's records deleted, they argue for the right to move such data to an alternative provider of the same service [9]. They cite similar historical examples in which Microsoft, IBM and Intel have been forced into making their systems interoperable with competitors, mainly through antitrust approaches in the USA and EU [8]. They argue that state intervention to break up these monopolies is not practical in a world dominated by competing national legislative regimes. Instead, they argue that merely providing users the ability to switch to alternatives would be sufficient. They believe that this would stimulate the development of service providers offering a range of alternative models [10].

The approach of treating monopoly service providers as public utilities is gaining ground in government circles. Recently the UK House of Lords called for the internet to be treated like a public utility rather than a market place of optional luxuries [43]. International bodies, such as the EU and UNESCO, have started calling for wider civic involvement in determining how services are provided [23,45] and for the development of alternative service provision models. For example, the outgoing EU Vice President, Neelie Kroes, stated in November 2014:

> "Why should we have to give up our privacy for a "free" service if we prefer to pay for that same service with cash and keep our privacy?" [23]

CONCLUSIONS - AGENCY

While these three dialectics focus on different issues, the poles of each axis rest on competing perspectives on the possibility of agency. Those who accept things as they are now do not see a possibility for agency, while their opponents do. On the first axis we have those who hold that preservation of privacy and delivery of service are necessarily in opposition. Here they are holding that there is no possibility of agency in the relationship between privacy and service design. To the contrary, we have seen how Langheinrich's design principles show multiple opportunities to intervene in the ways which deliver services while also maintaining privacy. In our second axis of ethical responsibility, the position of there being no ethical connection between the creator of an autonomous system and that system's effects is also a position of there being no agency. Here lack of agency pertains not to the nature of the system, but to the consequences of the system's actions. Under this view, once an autonomous system is activated, human agency ceases. However, as we have seen, preserving a lack of responsibility requires limiting the conception of where agency lies. Responsibility has to be defined in a very constricted manner which focuses on the making of each individual decision and denies the influence of any context. Instead, services are treated as independent of any human agency, in terms of their design, their environment, their purpose, how they are used and who benefits. By contrast, once autonomous services are contextualised within a field of human practice, human agency becomes apparent throughout the construction and operation of such systems and human ethical responsibility becomes self-evident. Finally, in our third axis of service architecture and business model, we see a historical lack of agency in the development of the broader internet culture. Here the client-server structure was accepted reflexively by developers and users, along with the most obvious reflections of this in organisational and economic models.

In all three debates we see one pole in each dialectic disempowering itself, primarily because it simply fails to recognise that there is a choice and that agency, the power to act differently, exists. The conclusion which emerges from this is that a key step to improving the current ethical status of cloud services is inculcating in programmers and leaders that they possess agency, bringing them to recognise there are alternatives and that they have the power to explore them.

REFERENCES

[1] Alessandro Acquisti, Allan Friedman, and Rahul Telang. 2006. Is there a cost to privacy breaches? An event study. *Proceedings of the Twenty-Seventh International Conference on Information Systems (citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.73.2942&rep=rep1&type=pdf.*

[2] Sadia Afroz, Aylin Caliskan Islam, Jordan Santell, Aaron Chapin, and Rachel Greenstadt. 2013. How Privacy Flaws Affect Consumer Perception. IEEE, 10–17. http://doi.org/10.1109/STAST.2013.13

[3] D.P. Anderson and G. Fedak. 2006. The Computational and Storage Potential of Volunteer Computing. *CCGRID '06 Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid*, IEEE Computer Society, 73–80. http://doi.org/10.1109/CCGRID.2006.101

[4] Katy Glenn Bass. 2013. *Chilling Effects: NSA Surveillance Drives U.S. Writers to Self-Censor*. PEN American Center. Retrieved from http://www.pen.org/sites/default/files/Chilling%20Effects_PEN%20American.pdf

[5] Lucas Bergkamp. 2002. The Privacy Fallacy. *Computer Law & Security Report* 18, 1, 31–47.

[6] Tim Berners-Lee. 1999. *Weaving the Web*. HarperCollins, New York, N.Y.

[7] Nathan Brooks. 2005. *Data Brokers: Background and Industry Overview*. Congressional Research Service, The Library of Congress, Washington, D.C.

[8] Ian Brown and Christopher T. Marsden. 2013. *Regulating code: good governance and better regulation in the information age*. The MIT Press, Cambridge, Mass.

[9] Ian Brown and Christopher T. Marsden. 2013. Regulating Code: Towards Prosumer Law? *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2224263

[10] Ian Brown and Christopher T. Marsden. 2013. Interoperability as a standard-based ICT competition remedy. *8th International Conference on Standardization and Innovation in Information Technology 2013*, IEEE, 1–8. http://doi.org/10.1109/SIIT.2013.6774570

[11] Comres. 2013. *Big Brother Watch Online Survey*. Comres, London.

[12] Eric P. Crampton and Donald J. Boudreaux. 2003. Does Cyberspace Need Antitrust? In *Who Rules the Net?*, Adam D Thierer and Clyde Wayne Jr. Crews (eds.). Cato Institute, Washington, D.C.

[13] Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. 2002. *The Platform for Privacy Preferences 1.0 (P3P1.0) Specification*. W3C. Retrieved from http://www.w3.org/TR/P3P/

[14] Mary J. Culnan. 1993. "How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use. *MIS Quarterly* 17, 3, pp. 341–363.

[15] James Curran, Natalie Fenton, and Des Freedman. 2012. *Misunderstanding the Internet*. Routledge, London.

[16] Tony Dyhouse. 2010. Addressing the silo mentality. *Infosecurity* 7, 2, 43. http://doi.org/10.1016/S1754-4548(10)70043-7

[17] Federal Trade Commission. 2014. *Data Brokers: A Call for Transparency*. Federal Trade Commission.

[18] Tom Hamburger and Matea Gold. 2014. Google, once disdainful of lobbying, now a master of Washington influence. *The Washington Post*. Retrieved June 23, 2015 from http://www.washingtonpost.com/politics/how-google-is-transforming-power-and-politicsgoogle-once-disdainful-of-lobbying-now-a-master-of-washington-influence/2014/04/12/51648b92-b4d3-11e3-8cb6-284052554d74_story.html

[19] Veikko Ikonen, Minni Kanerva, Panu Kouri, Bernd Stahl, and Kutoma Wakunuma. 2010. *D.1.2. Emerging Technologies Report*. ETICA Project.

[20] Deborah G. Johnson. 2011. Software Agents, Anticipatory Ethics, and Accountability. In *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight*, Gary E. Marchant, Braden R. Allenby and Joseph R. Herkert (eds.). Springer Netherlands, Dordrecht, 61–76. Retrieved February 17, 2015 from http://www.springerlink.com/index/10.1007/978-94-007-1356-7_5

[21] Aleksandra Korolova. 2011. Privacy Violations Using Microtargeted Ads: A Case Study. *Journal of Privacy and Confidentiality* 3, 1.

[22] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111, 24, 8788–8790. http://doi.org/10.1073/pnas.1320040111

[23] Neelie Kroes and Carl-Christian Buhr. 2014. Human society in a digital world. *Digital Minds for a New Europe*. Retrieved January 27, 2015 from https://ec.europa.eu/commission_2010-2014/kroes/en/content/human-society-digital-world-neelie-kroes-and-carl-christian-buhr

[24] Marc Langheinrich. 2001. Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. *Proceedings of the Third International Conference on Ubiquitous Computing*, Springer-Verlag, 273–291. Retrieved from http://www.vs.inf.ethz.ch/publ/papers/privacy-principles.pdf

[25] Alex Marthews and Catherine Tucker. 2014. Government Surveillance and Internet Search Behavior. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2412564

[26] Andreas Matthias. 2004. The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology* 6, 3, 175–183. http://doi.org/10.1007/s10676-004-3422-1

[27] Robert W. McChesney. 2014. Be Realistic, Demand the Impossible: Three Radically Democratic Internet Policies. *Critical Studies in Media Communication* 31, 2, 92–99. http://doi.org/10.1080/15295036.2014.913806

[28] Brendan Meeder, Jennifer Tam, Patrick Gage Kelley, and Lorrie Faith Cranor. 2010. RT@ IWantPrivacy: Widespread violation of privacy settings in the Twitter social network. *Proceedings of the Web*, 1–12.

[29] Peter Mell and Timothy Grance. 2011. *The NIST Definition of Cloud Computing*. National Institude for Standards & Technology.

[30] Seumas Miller. 2008. Collective Responsibility and Information and Communication Technology. In *Information Technology and Moral Philosophy*, Jeroen van den Hoven and John Weckert (eds.). Cambridge University Press, Cambridge; New York, 226–250.

[31] Eben Moglen. 2010. Freedom in the Cloud. Retrieved from http://www.softwarefreedom.org/events/2010/isoc-ny/FreedomInTheCloud-transcript.html

[32] OECD. 2013. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. OECD. Retrieved from http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

[33] OECD. 2013. *Exploring the Economics of Personal Data*. Retrieved June 23, 2015 from http://www.oecd-ilibrary.org/science-and-technology/exploring-the-economics-of-personal-data_5k486qtxldmq-en

[34] Office of the Data Protection Commissioner. 2015. *Annual Report of the Data Protection Commissioner of Ireland 2014*. Data Protection Commission of Ireland.

[35] Office of the Information & Privacy Commissioner of Ontario and IBM. 2011. *Privacy by Design: From Theory to Practice*. Office of the Information & Privacy Commissioner of Ontario, Ontario.

[36] Johan Pouwelse, Paweł Garbacki, Dick Epema, and Henk Sips. 2005. The Bittorrent P2P File-Sharing System: Measurements and Analysis. In *Peer-to-Peer Systems IV*, Miguel Castro and Robbert van Renesse (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 205–216. Retrieved June 23, 2015 from http://link.springer.com/10.1007/11558989_19

[37] RDF Working Group. 2015. RDF - Semantic Web Standards. *RDF - Semantic Web Standards*. Retrieved June 23, 2015 from http://www.w3.org/RDF/

[38] Paul Resnick, R. Kelly Garrett, Travis Kriplean, Sean A. Munson, and Natalie Jomini Stroud. 2013. Bursting your (filter) bubble: strategies for promoting diverse exposure. *Proceedings of the 2013 conference on computer supported cooperative work companion*, ACM Press, 95. http://doi.org/10.1145/2441955.2441981

[39] Robert Sparrow. 2007. Killer Robots. *Journal of Applied Philosophy* 24, 1, 62–77. http://doi.org/10.1111/j.1468-5930.2007.00346.x

[40] W. Richard Stevens and Gary R. Wright. 1994. *TCP/IP Illustrated Volume 1: The Protocols*. Addison-Wesley, Reading, Mass.

[41] Galen Strawson. 1994. The Impossibility of Moral Responsibility. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 75, 1/2, 5–24.

[42] Latanya Sweeney. 2013. Discrimination in Online Ad Delivery. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2208240

[43] The Select Committee on Digital Skills. 2015. *Make or Break: The Digital Future*. The Authority of the House of Lords, UK. Retrieved from http://www.publications.parliament.uk/pa/ld201415/ldselect/lddigital/111/111.pdf

[44] Joseph Turow. 2011. *The Daily You: How the New Advertising Industry is Defining Your Identity and Your World*. Yale University Press, New Haven.

[45] UNESCO Secretariat. 2014. *Internet Universality*. United Nations Educational, Scientific and Cultural Organization (UNESCO). Retrieved from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/news/internet_universality_en.pdf

[46] Verizon. 2014. *2014 Data Breach Investigations Report*. Verizon Enterprise Solutions.

[47] XML Working Group. 2015. XML Core Working Group Public Page. *XML Core Working Group Public Page*. Retrieved from http://www.w3.org/XML/Core/#Publications

[48] 2015. *Future-proofing Privacy*. Hogan Lovells, London.

[49] Privacy | Samsung UK. Retrieved June 18, 2015 from http://www.samsung.com/uk/info/privacy-SmartTV.html

[50] Amazon.com Help: Kindle Terms of Use. Retrieved June 22, 2015 from http://www.amazon.com/gp/help/customer/display.html?nodeId=200506200

Digital Alienation as the Foundation of Online Privacy Concerns

Brandt Dainow
Department of Computer Science, Maynooth University
Kildare, Co. Kildare
Ireland
+353 86 248 2846
brandt.dainow@nuim.ie

ABSTRACT

The term 'digital alienation' is used in critical IS research to refer to manifestations of alienation online. This paper explores the difficulties of using a traditional Marxist analysis to account for digital alienation. The problem is that the activity people undertake online does not look coerced or estranged from the creator's individuality, both of which are typically seen as necessary for the production of alienation. As a result of this apparent difficulty, much of the research has focused on the relationship between digital alienation and digital labour.

This paper attempts to overcome these difficulties by discarding the traditional approach. We argue one can better understand digital alienation by focusing on the relationship between user intent and technical infrastructure, rather than concerns with labour. Under the existing economic model dominating the internet, free services are financed by recording user activity and then using the products of this commercial surveillance to sell information about people to others. We show how the real harm in current online business models is that commercial surveillance is being used to commodify private life.

Seeking to define personal data in more precise terms, we will introduce two new concepts necessary for a detailed discussion of any ethical issues regarding personal data - the digital shadow and the digital persona. We will then show how affordances in current online systems are tuned to commodification of the user's personality. We will then explore the nature of online surveillance and show how affordances combine with the surveillance economy to produce digital alienation.

Categories and Subject Descriptors

K.4.1 **[Computers and Society]:** Public Policy Issues: Ethics, Privacy

General Terms

Management, Economics, Human Factors.

Keywords

digital alienation, privacy, digital economy, surveillance, targeted advertising, personalization, critical theory, ICT ethics, Marxism

1. INTRODUCTION

Digital alienation is a privacy issue. Digital alienation occurs when one's digital lifeworld or the digital self is exploited. The process of exploitation extracts value from a person's digital activity through coercion and manipulation. We are coerced into submission to ubiquitous commercial surveillance[1] of our digital activity. Value is extracted from this surveillance process through the conversion of surveilled data into economic and political capital. The entire system represents a reification of one's digital lifeworld and commodification of the digital self. It also poses a number of problems for traditional understandings of concepts related to alienation within Marxist theory, such as coercion, exploitation, and power dynamics. Indeed, much of the debate in this area over the last few years has been concerned with how to account for alienation within a digital context. I believe the solutions to current problems can best be achieved by altering the analytic approach.

2. SCOPE OF CONCERN

Being connected to the ubiquitous computing environment which is coming to surround us is already necessary for full participation in modern Western societies. A review across the range of those emerging ICT's which will impact society over the next decade shows that being connected may become necessary for survival itself [35]. When the internet first emerged, it was predicted that it would "flatten" the power structures of traditional society, even lead to the "fading away" of the nation state [57]. Such views were based on technological determinism; they envisioned the new distinguishing features of internet technology as passing unmodified into society and reshaping it to match the internet's technical architecture [15].

In reality, the development of the internet ecosystem has been filtered through the structures of pre-existing society and evolved in accordance with its imperatives. While it has been disruptive in terms of changing some of the dominant players in media markets, destroying some and creating others, it has not fundamentally changed the power structures in society. Authoritarian governments have learned to control and censor it, hegemonic corporate capitalism has come to dominate it, and people's digital activities have been cajoled into closed silos controlled by a very few exceptionally large corporations [15]. Once seen as the antidote to structural inequality, the internet has actually become a profoundly powerful tool of domination based on exploitation and alienation.

3. ALIENATION IN CRITICAL IS STUDIES

The term 'digital alienation' is used in Critical IS research to refer to manifestations of alienation online. Stemming from

[1] Commercial surveillance involves the recording and analysis of online user behaviour with the aim of predicting and controlling their behaviour [81].

digital labour studies [43] the focus soon bridged into social networking. A good example of this bridging can be seen in Fuchs and Sandoval's *Framework for Critically Theorising and Analysing Digital Labour* [25]. Initially exploring the dimensions of paid digital labour, the authors extend the analysis into the realm of unpaid labour within content production in social networks. P.J. Rey's paper *Alienation, Exploitation and Social Media* [66] explores the mechanisms by which capitalism has come to exploit social media. Rey's task involves demonstrating how alienation exists within social networking as a dynamic of value extraction. This approach is also used by Christian Fuchs [23,25] in most of his work. By contrast, Krüger and Johanssen's *Alienation and Digital Labour—A Depth-Hermeneutic Inquiry* [43] examines alienation through a survey of prosumer's comments about social network's themselves. Here alienation is demonstrated through the effects of the social network system's activities, rather than through the dynamics of labour and surplus value extraction. If alienation can derive from unpaid digital labour, as seen in social networks, the possibility arises that alienation can be found wherever unpaid digital labour occurs. Here we find Marc Andrejevic's *Surveillance and Alienation in the Online Economy* [6], which extends the analysis of alienation beyond social networking into general online activity. This paper shows a third approach to explaining digital alienation by focusing on exploitation, in contrast to the previously mentioned papers, which focus on value extraction and coercion. What all these analyses demonstrate is that the nature of alienation online necessarily diverges from the account of alienation in earlier, pre-digital, analyses. These divergences reflect the differences in structures of production and value-extraction between analogue and digital socio-technical systems. These differences are significant to the degree we may warrantably talk of a distinct "digital" form of alienation.

In Marx, alienation is the result of labour activity coerced into alienated forms in order to produce products estranged from the producer [60]. The political dynamic is the extraction of value from controlled and structured worker activity. Historically, analysis of digital alienation has focused on accounting for the traditional mechanisms underpinning alienation within a digital context. There has been an unspoken consensus that an account of digital alienation requires identifying the same structures and mechanisms within the digital context as Marx identified within the factory. Here the concern is to understand digital alienation by analysing it as the result of conditions considered necessary for alienation - coercion, labour and estrangement from product. With regard to coercion, the difficulty is whether people who freely choose to use social networks like Facebook can be described as coerced. The concern with labour is whether people's unpaid production of content in social networks can be described as labour. Finally, if people seem to be expressing themselves within social networks, the question arises as to how can they estranged from the output of their activity.

At one extreme researchers such as P.J. Rey have argued that the differences between the Victorian factory of Marx's analysis and modern digital activity are so great that alienation is a questionable concept within a digital context [66]. Rey argues that the products of digital labour in social networks are not alienated because creation of this content is freely chosen and creative. Referring back to Marx's categorisation of imagination as a distinguishing characteristic between animals and humans, Rey suggests that the creative nature of social content production renders the output unalienated. His view is that the creativity involved in social network content creation allows the producer to recognise themselves within their output. In addition, the free choice to engage in social networking means this labour is uncoerced. Rey does accept there is some degree of exploitation involved because social networks derive financial value from this output without financially compensating those who produced it. However, he argues this exploitation is mild because producers do receive compensation in other forms of capital. Rey argues that social network users are compensated because they retain use of their output for their own purposes. They can therefore use the content they produce to generate social and cultural capital. His position is that the non-economic value derived is so great that any exploitation is "relatively minimal" [66:415]. Furthermore, any exploitation present is, Rey argues, further diminished by the unalienated nature of prosumer output. Rey acknowledges that social networks also derive value from surveillance of user activity, usually without users being aware of it. However, while he sees this as mildly exploitative, he does not consider it alienating. Rey's position is that digital capitalism can maintain the inequalities and power structures within society identified by Marx, but without the need for alienation, or even very much exploitation.

It is notable that, while recognising that social networks extract value from user surveillance, Rey does not extend this recognition to the fact, noted by others, that such surveillance is almost universal throughout the internet [51,81]. A 2012 study of the world's busiest websites revealed that 94% engaged in some form of user surveillance themselves, half of whom also allowed unidentified third parties to engage in such tracking through their sites. The same study also found that 91% of these sites changed their content to match their understanding of the user [70], something impossible without a pre-existing knowledge of that user; knowledge which can only have come from previous surveillance. User activity in other parts of the internet, such as search and reading, does not generate cultural or social capital, but is still subject to the same levels of commercial surveillance. Following the logic of Rey's analysis, this renders such surveillance much more exploitative. In general, Rey's analysis treats technology as invisible and as permitting users to fully express themselves in an unmediated fashion. While Rey recognises that surveillance occurs, he fails to take into account that much it is used to tune and filter the online environment surrounding the user. Users are presented with "personalised" choices, links and content based on the results of covert surveillance as much as on the content they produce; something often referred to as the "filter bubble" [56,65]. People are therefore not able to make free choices or even fully express themselves, because the technology available to them is not value-neutral, but tuned to commodification [19,39].

Andrejevic's analysis of digital alienation is founded on just this consideration. All internet users are subject to pervasive universal surveillance by commercial enterprises [16,70,79,81]. The value of this surveillance far exceeds that derived from social network content creation [16,20,29,51,79,81]. Initially this information was used only to tune advertising delivery [16,81]. However, this information is now also used to tune the delivery of news on many sites [81] and for political manipulation [1]. Users have no choice over whether their activity online is recorded, processed and used, nor do they know who by [70]. This constitutes, for Andrejevic, alienation. His argument is that the lack of choice over whether to be surveilled or not constitutes a structurally-embedded coercion. He further argues that the lack of knowledge about this surveillance constitutes an epistemological alienation. Finally, he argues that the use of this information to alter content in an effort to manipulate the user fits Marx's definition of alienation as an estranged power structure working against the individual [6,8].

In contrast to Rey's position that exploitation is mild because the user derives non-economic use-value from the content they create, Christian Fuchs [23] has argued that exploitation is either present or not, and cannot be present in variable degrees. One cannot be a little bit exploited. Fuch's work tends to focus on the mechanisms of value-extraction within a digital context. Fuch's position is that any activity conducted by someone which can be used to generate economic value is labour. He seeks to bring together the competing positions held by Andrejevic and Rey by arguing that Rey is focused on subjective feelings of alienation whereas Andrejevic is focused on the objective conditions of non-control and non-ownership. However, Fuchs firmly comes down on the side of alienation being objectively present, arguing that the purported social use-value that content creators derive from their work hides the true commodity character of social networking [24]. He identifies two dimensions of value within social networks - the value of created content and the value of user presence. Here Fuchs agrees with Andrejevic that the users of social networks are themselves treated as commodified products which are then sold.

Both Fuchs and Andrejevic limit their conception of the use of personal data to the realm of advertising delivery. While the first use of this information was indeed to tune content, especially advertisements, to the user profile, this information is now also sold for other purposes, including political manipulation [1], credit scoring [54,72], housing and employment [12] and news delivery [81]. It is worth noting that both Facebook and the international trade body for online advertising, the Internet Advertising Bureau, agree with this assessment of where the real value lies in commercial online surveillance [16,20]. In comparison with this vast and pervasive surveillance industry, user-generated content within social networks is a trivial consideration. Under this analysis, alienation is a pervasive and unavoidable adjunct to almost all digital activity.

Rey, Andrejevic and Fuchs all approach alienation within a digital environment by focusing on Marx's mechanisms for its production and explaining how and where these mechanisms can be found online. While general commercial surveillance is mentioned, it is not really the central focus of their analysis, nor does it alter their approach. My position is that we can better account for digital alienation if we can liberate ourselves from the form of Marx's account. Marx provided an analysis of how alienation occurred within a particular historical and technological context. As we have seen from the above, we encounter problems if we assume that this is the only mechanism by which alienation can occur or that all of these traditional mechanisms are necessary. My argument is that the features of digital alienation are so different from traditional alienation that a new account is necessary.

4. HOW ALIENATION OCCURS ONLINE

In defining alienation, Marx considered two factors, the nature of alienation and the means by which it is produced. The nature of alienation is that the individual is disconnected from the products of their labour by property ownership rights; they are alienated from ownership of both the product and the means of producing it. This constitutes the material base of alienation and is the product of power relations governing the production process. Marx's account involved material coercion by controlling access to the means of survival so as to force people into alienating labour. Analysis of exploitation on the internet has been distracted by the apparent lack of coercion motivating online activity and by the appearance of self-expression in social networks. However, our analysis becomes less complicated if treat social networking within the broader context of pervasive digital surveillance. Here we recognise that, while content production in social networks is voluntary and can be self-expressive, it is just one type of action within the wider class of voluntary and self-serving digital activity which includes search, shopping, email, use of maps, health trackers, life loggers and other digital services, not to mention general web surfing. This is important because the range of digital activities will continue to spread until it permeates most of our environment [63]. Because of this it is essential to treat the current state of affairs as an intermediate process moving towards more ubiquitous computing. Our analysis must recognise that the political and economic structures which affect us within the current digital domain are on a trajectory to dominate our entire existence, offline as well as online. It is important, therefore, to recognise that the frame of analysis cannot limit itself to voluntary activity knowingly making use of digital services. The infrastructure being created now will one day support smart cities, the internet of things, and digital devices implanted within our bodies. Our entire existence will become mediated through digital services within a few decades [63].

Thus the place of labour as seen in a traditional account of alienation becomes problematic when value is being extracted from broad-spectrum use of digital services for life in general. Assuming that labour is a necessary precondition for alienation requires explaining how all activity using digital services constitutes labour despite the fact it generates no obvious income and may not even be anything more than a traditional activity, like walking or driving, which has been supplemented with a digital component. Certainly the argument of remuneration in the form of social or cultural capital is inapplicable with reference to activities which do not involve any form of communication, such as using search engines or passively reading a website, yet value is extracted from these activities by others via commercial surveillance [8,11,16,62,69,73,81]. If we redefine 'labour' as referring to any activity from which value may be drawn by any party, as Fuchs does [23,24], then almost all activity becomes labour and the term ceases to provide any real distinction from other mode of activity. I think it is better to abandon the issue of whether online activity is labour or not. There is nothing within Marx's description of alienation which requires that it must, of necessity, derive from labour. 'Alienation' in Marx is not a single concept, but a translation of two terms, *Entfremdung* and *Entäusserung*, which can also be translated as 'estrangement' and 'externalization' respectively [60]. These terms are applied to a variety of phenomena, including internal mental states, property relations and societal structures. It is true that Marx attempts to provide a systematic analysis of political economy based on the concept of alienated labour in his early work, but that attempt is incomplete [86]. In his later works, alienation becomes a descriptive term which is applied to multiple phenomena. There is nothing in his usage which locks alienation to labour except as a historically contingent feature of nineteenth century capitalism [86]. All that is required by Marx's account is that there be human activity and that this occur within certain types of unequal power structure within the field of economic competition.

On this basis, I propose to focus on digital alienation as a product of property relations regarding data. Surveillance is a process of data acquisition; some generated as the output of online surveillance monitoring systems and some data taken from elsewhere, such as the passenger name records used for international travel, geo-location data and credit scores [1]. The common element all these data elements have is that they are held to pertain to the same individual[2]. The dataset created is termed a "personal profile", as opposed to group profiles [31,38]. The personal profile is a digital representation of an individual. It is the central commodity of the surveillance economy. Each organisation which holds a personal profile subjects it to algorithmic analysis and manipulation in order to extract value from it. The term used in the industry is to "monetize" it. It is this profile which is used to tune content and for purposes of manipulation. All actions using personal data draw that data from the personal profile. Such use constitutes Marx's concept of an environment which reflects back on the producer estranged output

[2] This belief may be mistaken, it is not always possible to distinguish between a person and a device; and cases of mistaken identity also occur.

[8]. In that surveillance technology produces the personal profile as a commodity, it is a type of production process. The raw material for this production process is the activity of individuals [32], which is used to produce personal profiles. This production process is not owned by those who generate the activity which feeds it. This is the basis for alienation from ownership of the means of production. The surveillance process is hidden and unwelcome [14,32] and therefore represents an unequal, coercive and exploitative power structure [8]. We may view the personal profile as a field of contention between commercial surveillance companies and those who use their products on one hand opposed by individuals and privacy advocates on the other.

The essential starting point to all forms of alienation is individual activity. I therefore believe we may best understand digital alienation by examining the mechanisms by which an individual's digital activity is alienated. Here we must focus on the nature of personal action within a digital context, the mechanisms by which the personal profile is generated, and the use to which it is put. As mentioned above, the first task is to dispense with the need for a concept of labour. In Marx's analysis labour was the term used to distinguish activity which supported alienation from activity which did not. Labour supported alienation because it was activity which occurred within, and was shaped by, exploitative power structures. However, no such distinction between labour and non-labour exists online because all activity is surveilled and exploited [73,81]. Not only does discarding the need for labour ease our analysis, I believe it helps to direct our attention to the ubiquity of digital surveillance. Instead I will define human activity within a digital context in terms of people's intentions and expectations. To do this I will introduce distinguish the two targets of surveillance; communicative activity and everything else. I will refer to these as the 'digital persona' and the 'data shadow', respectively.

'Digital persona' is the term I propose for the body of digital material created by an individual through acts of online communication. The digital persona includes blogs, comments, product reviews, tweets and other social network postings, together with any other conscious communication by an individual within a digital context. Thus the digital persona is created by the individual to express and communicate. The digital persona is not a direct or unmediated reflection of the personality, but a creation through which the individual seeks to represent of an aspect of themselves. The disconnect between the offline and digital world permits people to exaggerate or repress particular aspects of their personality [77]. For example, introverts may use the digital persona to compensate for difficulties they have in face-to-face interactions [3] while extroverts often use it to confirm pre-existing characteristics [82]. In other cases, people develop new personal characteristics online so that they can incorporate them into their offline personality [53]. In all cases, what is revealed or portrayed is further influenced by previous experiences online, especially concerns over privacy and security [42,87]. I derive the term 'persona' from C. G. Jung's concept of the persona as a creation of the ego designed to represent a subset of that ego within specific social circumstances [36]. The same idea is used within a sociological perspective in Goffman's *The Presentation of Self in Everyday Life* under the term 'masks' [28], which outlines his Dramaturgical Theory, a sub-set of Symbolic Interactionism [68].

The term 'data shadow' was first used by Alan Westin in *Privacy and Freedom* [84], but has entered into general use both in computing and privacy discussions. It refers to the information generated by someone as a side-effect of their use of digital technology. These days this includes log files, access records, search histories, movements between and within web sites, mobile phone location records and all financial activities not involving cash [11,31,39]. Thus the term 'data shadow' refers to all digital information pertaining to an individual which they did not consciously and intentionally create for communicative purposes. This information may have been generated by the user for other purposes, such as their "click-stream history," which is a record of their mouse click activity within a website [32], or their "search history," a record of all the searches they have made in a given search engine. Elements of the data shadow can also be generated through the monitoring and recording of user activity by other systems. For example, web server log files, containing records of every file request, constitute data generated by the system about the user. The term 'data shadow' includes the material used to commodify users within social networks, but also applies outside social networking. Data shadows may be created through any and all use of digital technology.

Data shadows are created by a network of commercial surveillance agencies whose tracking technologies permeate digital services [11,16,46,51,81]. Very few of these agencies are known to the public [11,73]. Some, like Google and Facebook, are well known because of their public profile as digital service providers, though their activity as commercial surveillance agents is less well known, even though it drives their profits [20,29]. Others, such as DoubleClick, Acxiom, Experian and BlueKai are known to industry analysts and privacy advocates as a result of their scale and reach. However, the majority, such as ClickTale, Optimzly, Kiss Metrics, Info Group, Ace Metrics, Crazy Egg, Site Meter, Moz, Adgistics, People Metrics, Data Dog, Data Mentors, Extrawatch, Inspectlet, eDataSource, Prognoz, and literally hundreds of others, are unknown outside the specialist profiling industry. No one knows how many of these agencies there are, or what they do, but it is known they combine the data they gather with information from other sources to create detailed profiles on literally hundreds of millions, if not billions, of people [11,21,83]. The commercial surveillance industry is much larger in terms of economic value and user-base than any other online industry [16,73,81].

This universal commercial surveillance means there is no way to use most digital services without being surveilled [21,32,73,81,83]. For most digital services there is no alternative provider who does not practice surveillance (or permit others to do so) within the service stream [76,81]. However, lack of choice most strongly stems from lack of knowledge. We are simply unaware of when we are being surveilled, who by and for what purpose [32,79]. Obviously, one cannot exercise choice over things one is unaware of. As we have seen, this lack of choice has been held to constitute coercion by Andrejevic and Fuchs, but not by Rey. Lack of choice as coercion has a long history of support in philosophy. For example, Aquinas argues that coercion occurs when actions by one person mean someone cannot act otherwise [10]. However, this position was challenged in the twentieth century by the position that coercion requires communication between the coercer and their target, usually in the form of conditional threats [2]. Under this view coercion is a communicative act, not a contextualising situation. This is the position currently supported in much legal practice, especially in the USA [5]. However, since the 1980's arguments have re-emerged in support of structural coercion; the creation of situations in which one is prevented from selecting alternative courses of action [67]. Here the focus is shifted to the coercer's intentions to remove choice from another [4]. This accords with much of Marx's analysis in which he focuses on the general circumstances of capitalist society as coercive in the sense of removing freedom [86]. Clearly, hiding surveillance so that people cannot avoid it constitutes removal of choice and diminution of freedom. Thus it is possible to argue from this perspective that the lack of choice to avoid surveillance constitutes coercion. However, we must recognise that this position is not in accord with how many, especially in jurisprudence, understand the term.

Lack of choice, even coercion, does not automatically mean that the output of a productive process is alienated. I wish therefore to explore the mechanism by which digital activity becomes alienated. Since we have two forms of digital data, the digital persona and the data shadow, two accounts are necessary. I shall commence with the alienation of the digital persona.

5. EGO, AFFORDANCES AND THE DIGITAL PERSONA

People use Web 2.0 technologies to create their digital persona. The process by which they do this, and the persona they create, are alienated. We therefore need an account of the mechanism by which people do this and how alienation occurs. Central to my account of how the digital persona is alienated is the view of technology as a socio-technical system [35]. A technology may be composed of multiple artefacts and may be "read" or understood in different ways [30,34]. The nature of the "reading" depends on the person, their social environment, past experiences and other factors, all of which are constrained by the functional capabilities of the artefacts in question [58]. We therefore need a conceptual framework which holds all the dynamics which are at play in a person's understanding and use of a technical system. I will use the concept of "affordances" to explain the interaction between people and the technical artefacts.

The concept of affordances originates with James Gibson's conceptualisation on the subject of how animals perceive and understand their environment in *The Ecological Approach to Visual Perception* [27]

> "The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill… It implies the complementarity of the animal and the environment… [affordances] have to be measured relative to the animal. They are unique for that animal. They are not just abstract physical properties." [27:127]

Gibson was arguing against a reductionist understanding of perception and for a perceptive process within all animals in which perception itself is not merely a process of physical activity onto which understanding is overlaid *post hoc*. Instead, he argued that the perceptive process itself incorporates cognitive elements such as motivation, environmental context and past experience into the act of seeing.

The concept was applied to ICT analysis by Ian Hutchby in *Technologies, Texts and Affordances* [34], in which he describes technologies as

> "Texts which are written in certain ways by their developers, producers and marketers, and have to be read by their users or consumers. The writers of these technology texts may seek to impose particular meanings on the artefact, and to constrain the range of possible interpretations open to users. Users, by contrast, may seek to produce readings of the technology texts which best suit the purposes they have in mind for the artefact… Neither the writing or reading of technology texts is determinate: both are open, negotiated processes. Although there may be ways that technology texts have preferred readings built into them, it is always open to the user to find a way around this attempt at interpretive closure." [34:445]

We may thus see affordances as a field of competition in which the owners of a technology compete with the users of that technology for domination of the affordances dictating how that technology is understood and used. Donald Norman explores this competition over technological affordances in *The Design of Everyday Things* [58]. In Norman's account, we use affordances to build conceptual models of how things work. Any technology involves the interaction of two conceptual models; a design model and a user model. The design model is the conceptual model held by the designers when they built the technology and in accord with which they try to construct the artefact. The user's model is the conceptual model users have of that same technology. Norman is concerned with what happens when the two models clash or diverge. According to Norman, there is no necessary convergence between the user's mental model and the designer's. In fact, in Norman's view, the two model's clash most of the time. Using Norman's framework, I suggest that the user model conceptualises the Web 2.0 services people use to express their digital personas as private, unmediated and natural. The user model fails to recognise the degree of surveillance and the degree to which their activities are mediated through a technology designed for data gathering and commodification. Users also fail to recognise the degree to which surveillance is used to filter and control the content they see in social networking and news sites and in advertising. Instead, users see the content presented to them within social networks as somehow neutral, unmediated and unsurveilled [71]. In contrast, service providers, such as Google and Facebook, show evidence of believing that users have the same conceptual model as designers. They have countered concerns over online privacy by stating users have no expectation of privacy and accept that the material they create will be processed for purposes of commodification [22].

There are numerous studies which demonstrate that users manipulate their self-expression online in order to convey specific characteristics and control the image others have of them [41,50,53,82]. In our terminology we may say people use Web 2.0 technologies to construct their digital personas. Their understanding of what can be expressed, the values determining what should be expressed and how this is to be done are determined by the affordances users perceive in these technologies [74,89]. These affordances constitute what Groffman describes as the "props and tasks" [28:143] which dictate what persona[3] is appropriate and the "expressive resources" [55:74] available from which to construct it.

Unfortunately for users, Facebook and similar Web 2.0 systems are not designed for people to portray themselves in any manner they may choose. Instead, Facebook and similar systems divide personal characteristics into a set of discrete data points, such as preferred objects of consumption, marketable skills, and approvable attitudes [33]. Furthermore, qualitative characteristics, such as friendship, are reduced to quantitative values, such as the number of likes or followers. Facebook's affordances, in particular, suggest to users that their digital persona is a true reflection of their identity, yet is at the same time something to be constructed, managed and enhanced [26]. Facebook openly expresses the neo-liberal concept of a "personal brand," in which a person creates a commodified public image as the repository of their social capital [44]. The affordances of Facebook present the individual as composed of consumption patterns (such as preferred movies, books and music) and patterns of association (as shown through one's likes, friends and photos). These are dimensions of analysis more suited to processing for advertising than developing an understanding of the whole person. There is good empirical evidence that this model conflicts with the affordances the user brings to Facebook. In many cases users seek to express themselves in ways restricted by the affordances Facebook imposes, resulting in dissatisfaction, resistance and disuse [26,42,50,74].

In using affordances tuned to atomising, quantifying and commodifying the depiction of people, social media systems like Facebook alienate the digital persona. Rather than a free expression of the self, users are forced to display only those

[3] Groffman's term is "performance" [28:143]

characteristics which are commodifiable. These characteristics are then embedded in a manipulated content environment which reinforces and promotes ongoing commodification, and therefore embeds the alienated digital persona within an alienated social environment.

6. ALIENATION AND THE DATA SHADOW

The mechanisms by which the data shadow is alienated are straightforward compared to the digital persona and commence with unavoidable, hidden, ubiquitous commercial surveillance [70,73,81]. In that the digital world is permeated with unknown entities gathering unknown information to use for unknown purposes [21,83], the digital environment is self-evidently epistemologically alienated from the user. The material base of the commercial surveillance system supports a superstructure devoted to exerting power over the individual by influencing their behaviour directly, or by influencing decisions made about them by other people [11,72,80]. This is achieved through personalization of content [56], such as the advertising [78] and news [81] to which people are exposed.

The unavoidability of commercial surveillance is made possible by the lack of ownership or control users have over digital services. It is known that, in general, people do not like commercial surveillance or content personalization [14,17]. Commercial surveillance therefore constitutes the exercise of power over individuals and a diminution of their freedom, another manifestation of alienation [7]. Furthermore, the knowledge that unknown surveillance is occurring, in combination with lack of knowledge about how that information is used, has a chilling effect on people's online activities [32,49,75]. In effect, people are alienated from their own actions online before they perform them. In that this chilling effect also applies to how people communicate online, ubiquitous commercial surveillance further alienates people from each other.

7. DIGITAL ALIENATION – THE COMPLETE PICTURE

We are now in a position to provide an account of how the four dimensions of alienation occur. First, users are alienated from their productive activity through restricted affordances within expressive Web 2.0 technologies which promote a commodity fetishism of personal characteristics and interpersonal relationships. This is made possible by an alienated power structure which is designed around treating users as commodities [8,23]. Users are alienated from non-expressive activity by the presence of ubiquitous hidden surveillance systems. Thus users are alienated from all forms of digital activity. Second, users are alienated from the products of their digital activity by property relations. These grant service owners the right to reuse user-produced content for their own purposes and to process both the digital persona and the data shadow in order to construct personal profiles. Users are further alienated from the products of their own activity since the personal profile is used against them, either to manipulate their behaviour or to influence how others treat them. In addition, the abandonment of the open standards which created the web means that the products of user activity are imprisoned within data silos owned by service providers [18]. Thus, you may close your Facebook account, but you can't move it to another social network. Third, users are alienated from each other by the necessary mediation of fetishizing social networks and by the chilling effect of ubiquitous surveillance. Finally, users are alienated from themselves and their own human potential in three ways; through the imposition of fetishizing affordances promoting the concept of the personal brand, through their limited control over their own digital persona, and through the use of personalization technologies which confine the user's ability to discover the unexpected, the unusual, and the uncommodified.

8. SOLUTIONS

No solution exists today which can resist these patterns and structures of digital alienation. However, a number of technologies exist which can form part of a solution, while the design principles to complete the solution are understood. Two related characteristics support the existence of digital alienation, lack of choice and lack of power. The solution is therefore to restore choice and empower the user. In my view solutions that look to regulation, such as data protection and privacy laws, merely perpetuate a hierarchical structure which keeps people in a powerless position. Instead of companies deciding what to surveil, we merely pass the decision to legislators. Given the history of government digital surveillance [9] there is nothing to suggest this improves matters. In addition, the impossibility of a single legislative framework for the entire internet [64,85,88] means surveillance companies can simply move to more conducive regimes. Furthermore, centralised storage of personal data is frequently subject to leaks [40,47,48], so I am opposed to centralised storage of any fashion, never mind under what rules.

The first task in combating alienation must be to remove coercion from the situation by giving users the choice over whether to be surveilled and for what purpose. A number of technologies exist which can offer elements of this solution. Anonymizing systems such as such as TOR [52] and TextSecure [61] enable users to avoid being tracked while using the existing internet. These need to be extended and built into a comprehensive set of easy-to-use systems which can wrap browsers and other applications in a protective and intelligent layer which negotiates and controls what data is accessed by what services. Protocols like the W3C's Platform for Privacy Preferences (PPP) [13] can form the basis for such communications. Design of data gathering systems should follow principles of privacy preservation, such as those developed by Marc Langheinrich [45], one of the authors of PPP. Such technology would enable users to control how much information is gathered about them and thus how much personalization is possible. This de-alienates the productive technology by putting control in the hands of the user and de-alienates their digital environment by permitting them to control or prevent personalization.

While these solutions restore choice to the user, they only partially redress the balance in an existing system which is structurally inequitable. The long-term solution must therefore be to move personal data storage, and therefore ownership, into the hands of the users. Here the solution is to reverse the cloud architecture. Currently, centralised systems run analyses of locally held data. I propose inverting this structure, such that personal data is held by the person in their own devices. Effectively each person, or home, would operate their own data store. Following Langheinrich's privacy-preserving design principles [45], devices would, wherever feasible, store their own data. A personal server or gateway would provide the interface between digital service providers and the user's personal data. This gateway would be able to negotiate access for services and prepare personal data for access. This pre-processing would anonymise the data to the degree selected by the user for that type of service. I envision this system working in a manner similar to hierarchical protection domains (or "security rings") within chipsets. These create a series of layers within which particular software operations can be confined so as to shield the system from inappropriate operations [37]. Corporate digital services could still be centrally managed and owned, but their computations would have to call on the individual's own data store rather than house it on corporate servers.

This is, however, merely a collection of artefacts. As a socially-embedded system, technology needs more than just hardware if it is to be adopted. The additional component required is therefore societal structures promoting and maintaining such a system. A network of local technicians is required to maintain and develop such systems, provide advice and training, lobby regulators for support and so forth. Here I suggest the basis lies in recognising the value of personal data. For example, the value of a Facebook user is between $US40 and $US300 [59]. If personal data has value for service providers, let them pay for it. A system of micropayments for access to personal data would create a data economy enabling individuals to earn money through the gathering and storing of their own data. Support agencies, such as technical staff and software vendors, can then be remunerated through a share of this income. Such a system would permit the development of an intermediate layer of data vendors who can store and provide personal data on the user's behalf, according to guidelines provided by those users, or remotely maintain data held in the home. Such a system permits of multiple organisational models. Community groups could operate such services. For example, people who share the same set of data access protocols could form cooperatives to manage storage and access to their member's data. As yet, such technology does not exist. However, the hardware is already in place. Personal cloud storage devices have been available for several years. These permit users to store their data in their home while still being able to access it remotely. The missing components are therefore the micropayment and data negotiation systems. Protocols exist which can handle both, they merely need to be implemented as working products.

We need to bear in mind that the digital service infrastructure we see today is merely a step towards a digital environment of ubiquitous devices; embedded within our bodies, throughout our homes, offices, cars and public spaces. A critical evaluation of current data practices must consider this long-term future and seek emancipatory paths within it. As we have seen, digital alienation is the product primarily of inequitable power structures which intentionally deny users control, or even knowledge, of what is being done to them. The motive power of these structures is the economic value of personal data. If digital services are to align with individual needs, we cannot avoid personal data being processed. The solution is therefore to develop systems which pass some of that value back to the user. Doing so gives the user power and makes them a viable partner for other organisations who can earn a living by controlling access to personal data on behalf of the user. Giving the individual control over their personal data emancipates them from subjection to hegemonic digital capitalism by permitting them to negotiate the terms of the relationship they have with their digital service providers.

8. REFERENCES

[1] Nathan Abse. 2012. *Microtargeted Political Advertising in Election 2012*. Internet Advertising Bureau, New York, NY, USA. Retrieved from http://www.iab.net/media/file/Innovations_In_Web_Marketing_and_Advertising_delivery.pdf

[2] Timo Airaksinen. 1988. An Analysis of Coercion. *Journal of Peace Research* 25, 3, pp. 213–227.

[3] Yair Amichai-Hamburger, Galit Wainapel, and Shaul Fox. 2002. "On the Internet No One Knows I'm an Introvert": Extroversion, Neuroticism, and Internet Interaction. *CyberPsychology & Behavior* 5, 2, 125–128. http://doi.org/10.1089/109493102753770507

[4] Scott Anderson. 2008. Of Theories of Coercion, Two Axes, and the Importance of the Coercer. *Journal of Moral Philosophy* 5, 3, 394–422. http://doi.org/10.1163/174552408X369736

[5] Scott Anderson. 2010. The Enforcement Approach to Coercion. *Journal of Ethics and Social Philosophy* 5, 1.

[6] Mark B. Andrejevic. 2011. Surveillance and Alienation in the Online Economy. *Surveillance & Society* 8, 3, 278–287.

[7] Mark B. Andrejevic. 2011. Estrangement 2.0. *World Picture* 6, 1–14.

[8] Mark B. Andrejevic. 2012. Exploitation in the Data Mine. In *Internet and Surveillance*, Christian Fuchs, Marisol Sandoval, Boersma Kees and Anders Albrechtslund (eds.). Routledge, New York, N.Y, 71–88.

[9] Julia Angwin. 2014. Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance. Henry Holt & Co, New York, N.Y.

[10] Thomas Aquinas. 1920. *Summa Theologica*. Fathers of the English Dominican Province, London.

[11] Nathan Brooks. 2005. *Data Brokers: Background and Industry Overview*. Congressional Research Service, The Library of Congress, Washington, D.C.

[12] Danielle Keats Citron and Frank Pasquale. 2014. The Scored Society. *Washington Law Review* 89, 1, 1–33.

[13] Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. 2002. *The Platform for Privacy Preferences 1.0 (P3P1.0) Specification*. W3C. Retrieved from http://www.w3.org/TR/P3P/

[14] Mary J. Culnan. 1993. "How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use. *MIS Quarterly* 17, 3, pp. 341–363.

[15] James Curran, Natalie Fenton, and Des Freedman. 2012. *Misunderstanding the Internet*. Routledge, London.

[16] John Deighton and Lenora Kornfield. 2012. *Economic Value of an Advertising-supported Internet Ecosystem*. Internet Advertising Bureau, New York, N.Y. Retrieved from http://www.iab.net/insights_research/industry_data_and_landscape/economicvalue

[17] Jenny van Doorn and JannyC. Hoekstra. 2013. Customization of Online Advertising: The Role of Intrusiveness. *Marketing Letters* 24, 4, 339–351. http://doi.org/10.1007/s11002-012-9222-1

[18] Tony Dyhouse. 2010. Addressing the Silo Mentality. *Infosecurity* 7, 2, 43. http://doi.org/10.1016/S1754-4548(10)70043-7

[19] Fredrik Erlandsson, Martin Boldt, and Henric Johnson. 2012. Privacy Threats Related to User Profiling in Online Social Networks. *Proceedings of 2012 International Conference on Social Computing*, IEEE, 838–842.

[20] Facebook Inc. 2014. *Facebook Inc. First Quarter 2014 Results*. Facebook Inc. Retrieved from http://investor.fb.com/releasedetail.cfm?ReleaseID=842071

[21] Federal Trade Commission. 2014. *Data Brokers: A Call for Transparency*. Federal Trade Commission.

[22] Jeremy Fogel. 2014. A Reasonable Expectation of Privacy. *Litigation* 40, 4. Retrieved December 15, 2014 from http://www.americanbar.org/publications/litigation_journal/2013-14/spring/a_reasonable_expectation_privacy.html

[23] Christian Fuchs. 2013. Class and Exploitation on the Internet. In *Digital labor - the Internet as playground and factory*, Trebor Scholz (ed.). Routledge, New York, 211–224.

[24] Christian Fuchs. 2014. *Digital Labour and Karl Marx*. Routledge, Taylor & Francis Group, New York, NY.

[25] Christian Fuchs and Marisol Sandoval. 2014. Digital Workers of the World Unite! A Framework for Critically Theorising and Analysing Digital Labour. *tripleC* 12, 2, 486–563.

[26] Ilana Gershon. 2011. Un-Friend My Heart: Facebook, Promiscuity, and Heartbreak in a Neoliberal Age. *Anthropological Quarterly* 84, 4, 865–894.

[27] James Jerome Gibson. 1986. *The Ecological Approach to Visual Perception*. Psychology Press, New York.

[28] Erving Goffman. 1990. *The Presentation of Self in Everyday Life*. Doubleday, New York, NY.

[29] Google Inc. 2014. *Google Inc. First Quarter 2014 Results*. Google Inc. Retrieved from http://investor.google.com/earnings/2014/Q1_google_earnings.html

[30] Martin Heidegger. 1977. The Question Concerning Technology, and Other Essays. Harper & Row, New York.

[31] M Hildebrandt. 2008. Defining Profiling: A New Type of Knowledge? In *Profiling the European Citizen*, M Hildebrandt and Serge Gurtwith (eds.). Springer, New York, NY.

[32] Simone van der Hof and Corien Prins. 2008. Personalisation and its Influence on Identities, Behaviour and Social Values. In *Profiling the European Citizen*, M Hildebrandt and Serge Gurtwith (eds.). Springer, New York, N.Y, 111–127.

[33] Gordon Hull, Heather Richter Lipford, and Celine Latulipe. 2011. Contextual Gaps: Privacy Issues on Facebook. *Ethics and Information Technology* 13, 4, 289–302. http://doi.org/10.1007/s10676-010-9224-8

[34] Ian Hutchby. 2001. Technologies, Texts and Affordances. *Sociology* 35, 2, 441–456. http://doi.org/10.1177/S0038038501000219

[35] Veikko Ikonen, Minni Kanerva, Panu Kouri, Bernd Stahl, and Kutoma Wakunuma. 2010. *D.1.2. Emerging Technologies Report*. ETICA Project.

[36] C. G. Jung. 1977. The Collected Works of C. G. Jung. 6: Psychological types. Pantheon Books, New York, NY.

[37] Paul A. Karger and Andrew J. Herbert. 1984. An Augmented Capability Architecture to Support Lattice Security and Traceability of Access. IEEE, 2–2. http://doi.org/10.1109/SP.1984.10001

[38] Helen Kennedy. 2008. New Media's Potential for Personalization. *Information, Communication & Society* 11, 3, 307–325. http://doi.org/10.1080/13691180802025293

[39] Byungwan Koh. 2011. *User Profiling in Online Marketplaces and Security*. ProQuest LLC, Ann Arbor, MI.

[40] Bert-Jaap Koops and Ronald Leenes. 2014. Privacy Regulation Cannot be Hardcoded. A Critical Comment on the "Privacy by Design" Provision in Data-Protection Law. *International Review of Law, Computers & Technology* 28, 2, 159–171. http://doi.org/10.1080/13600869.2013.801589

[41] Nicole C. Krämer and Stephan Winter. 2008. Impression Management 2.0: The Relationship of Self-Esteem, Extraversion, Self-Efficacy, and Self-Presentation Within Social Networking Sites. *Journal of Media Psychology* 20, 3, 106–116. http://doi.org/10.1027/1864-1105.20.3.106

[42] Hanna Krasnova, Oliver Günther, Sarah Spiekermann, and Ksenia Koroleva. 2009. Privacy Concerns and Identity in Online Social Networks. *Identity in the Information Society* 2, 1, 39–63. http://doi.org/10.1007/s12394-009-0019-1

[43] Steffan Krüger and Jacob Johanssen. 2014. Alienation and Digital Labour—A Depth-Hermeneutic Inquiry into Online Commodification and the Unconscious. *tripleC* 12, 2, 632–645.

[44] D. J. Lair. 2005. Marketization and the Recasting of the Professional Self: The Rhetoric and Ethics of Personal Branding. *Management Communication Quarterly* 18, 3, 307–343. http://doi.org/10.1177/0893318904270744

[45] Marc Langheinrich. 2001. Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. *Proceedings of the Third International Conference on Ubiquitous Computing*, Springer-Verlag, 273–291. Retrieved from http://www.vs.inf.ethz.ch/publ/papers/privacy-principles.pdf

[46] David Lazarus. 2015. Verizon's Super New Way to Mess with Your Privacy. *The Los Angeles times*.

[47] Yabing Liu, Krishna P Gummadi, Balachander Krishnamurthy, and Alan Mislove. 2011. Analyzing Facebook Privacy Settings: User Expectations vs. Reality. *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement*, ACM, 61–70.

[48] Michelle Madejski, Maritza Johnson, and Steven M Bellovin. 2012. A Study of Privacy Settings Errors in an Online Social Network. *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, IEEE, 340–345.

[49] Alex Marthews and Catherine Tucker. 2014. Government Surveillance and Internet Search Behavior. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2412564

[50] Silva Martin and Stef Nicovich. 2013. Self Concept Clarity and its Impact on the Self-Avatar Relationship in a Mediated Environment. *Atlantic Marketing Association Conference Proceedings*.

[51] Viktor Mayer-Schönberger. 2009. *Delete: the Virtue of Forgetting in the Digital Age*. Princeton Univ. Press, Princeton, NJ.

[52] Damon McCoy, Kevin Bauer, Dirk Grunwald, Tadayoshi Kohno, and Douglas Sicker. 2008. Shining Light in Dark Places: Understanding the Tor network. *Privacy Enhancing Technologies*, Springer, 63–76.

[53] Soraya Mehdizadeh. 2010. Self-Presentation 2.0: Narcissism and Self-Esteem on Facebook. *Cyberpsychology, Behavior, and Social Networking* 13, 4, 357–364. http://doi.org/10.1089/cyber.2009.0257

[54] Edmund Mierzwinski and Jeffrey Chester. 2013. Selling Consumers Not Lists: The New World of Digital Decision-Making and the Role of the Fair Credit Reporting Act. *Suffolk University Law Review* 46, 3, 855–867.

[55] Hugh Miller and Jill Arnold. 2001. Self in Web Home Pages. In *Towards CyberPsychology*, Giuseppe Riva and Carlo Galimberti (eds.). IOS Press, Oxford, 73–94.

[56] Sayooran Nagulendra and Julita Vassileva. 2014. Understanding and Controlling the Filter Bubble through Interactive Visualization: a User Study. *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, ACM Press, 107–115. http://doi.org/10.1145/2631775.2631811

[57] Nicholas Negroponte. 1996. *Being Digital*. Vintage Books, New York, NY.

[58] Donald A. Norman. 2002. *The Design of Everyday Things*. Doubleday, New York.

[59] OECD. 2013. *Exploring the Economics of Personal Data*. Retrieved June 23, 2015 from http://www.oecd-ilibrary.org/science-and-technology/exploring-the-economics-of-personal-data_5k486qtxldmq-en

[60] Bertell Ollman. 2001. *Alienation: Marx's Concept of Man in Capitalist Society*. Cambridge University Press, Cambridge; New York.

[61] Rolf Oppliger (ed.). 2014. *Secure Messaging on the Internet*. Artech House, Boston.

[62] Karl Palmås. 2011. Panspectric Surveillance and the Contemporary Corporation. *Surveillance & Society* 8, 3, 338–354.

[63] Michael Rader, A. Antener, Rafael Capurro, Michael Nagenborg, and L. Stengel. 2010. *D.3.2. Evaluation Report*. ETICA Project.

[64] Joel Reidenberg. 2005. Technology & Internet Jurisdiction. *University of Pennsylvania Law Review* 153, 1951–1974.

[65] Paul Resnick, R. Kelly Garrett, Travis Kriplean, Sean A. Munson, and Natalie Jomini Stroud. 2013. Bursting Your (Filter) Bubble: Strategies for Promoting Diverse Exposure. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*, ACM Press, 95. http://doi.org/10.1145/2441955.2441981

[66] P.J. Rey. 2012. Alienation, Exploitation, and Social Media. *American Behavioral Scientist* 56, 4, 399–420. http://doi.org/10.1177/0002764211429367

[67] Jan-Willem van der Rijt. 2012. *The Importance of Assent: a Theory of Coercion and Dignity*. Springer, Dordrecht.

[68] George Ritzer. 2011. *Sociological Theory*. McGraw-Hill, New York.

[69] Ira S. Rubinstein. 2013. Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law* 3, 2, 74–87.

[70] Marisol Sandoval. 2012. A Critical Empirical Case Study of Consumer Surveillance on Web 2.0. In *Internet and Surveillance*, Christian Fuchs, Marisol Sandoval, Boersma Kees and Anders Albrechtslund (eds.). Routledge, New York, N.Y, 147–169.

[71] Kellyton dos Santos Brito, Frederico Araujo Durao, Vinicius Cardoso Garcia, and Silvio Romero de Lemos Meira. 2013. How People Care About Their Personal Data Released on Social Media. *Proceedings of 2013 Eleventh Annual International Conference on Privacy, Security and Trust (PST)*, IEEE, 111–118. http://doi.org/10.1109/PST.2013.6596044

[72] Amy J Schmitz. 2015. Secret Consumer Scores and Segmentations: Separating "Haves" from "Have-Nots." *Michigan State Law Review* 2014, 5, 1411.

[73] Bruce Schneier. 2015. *Data and Goliath: the Hidden Battles to Capture Your Data and Control Your World*. Norton, New York, NY.

[74] Richard C. Sherman. 2001. The Mind's Eye in Cyberspace: Online Perceptions of Self and Others. In *Towards CyberPsychology*, Giuseppe Riva and Carlo Galimberti (eds.). IOS Press, Oxford, 73–72.

[75] Daniel J. Solove. 2004. *The Digital Person: Technology and Privacy in the Information Age*. New York Univ. Press, New York, NY [u.a.].

[76] Inger L. Stole. 2014. Persistent Pursuit of Personal Information: A Historical Perspective on Digital Advertising Strategies. *Critical Studies in Media Communication* 31, 2, 129–133. http://doi.org/10.1080/15295036.2014.921319

[77] John Suler. 2004. The Online Disinhibition Effect. *CyberPsychology & Behavior* 7, 3, 321–326. http://doi.org/10.1089/1094931041291295

[78] Latanya Sweeney. 2013. Discrimination in Online Ad Delivery. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2208240

[79] Omar Tene and Jules Polonetsky. 2012. To Track or "Do Not Track": Advancing Transparency and Individual Control in Online Behavioral Advertising. *Minnesota Journal of Law, Science & Technology* 13, 1, 281–358.

[80] Omer Tene and Jules Polonetsky. 2013. Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property* 11, 5, 239–272.

[81] Joseph Turow. 2011. *The Daily You: How the New Advertising Industry is Defining Your Identity and Your World.* Yale University Press, New Haven.

[82] Shaojung Sharon Wang. 2013. "I Share, Therefore I Am": Personality Traits, Life Satisfaction, and Facebook Check-Ins. *Cyberpsychology, Behavior, and Social Networking* 16, 12, 870–877. http://doi.org/10.1089/cyber.2012.0395

[83] Logan Danielle Wayne. 2012. The Data-Broker Threat. *Journal of Criminal Law and Criminology* 102, 1, 253–282.

[84] Alan F Westin. 1970. *Privacy and Freedom*. Bodley Head, London.

[85] S. Wilske and T. Schiller. 1997. International Jurisdiction in Cyberspace: Which States may Regulate the Internet? *Federal Communications Law Journal* 50, 119 – 178.

[86] Allen W. Wood. 2006. *Karl Marx*. Routledge, New York.

[87] Mike Z. Yao and Daniel G. Linz. 2008. Predicting Self-Protections of Online Privacy. *CyberPsychology & Behavior* 11, 5, 615–617. http://doi.org/10.1089/cpb.2007.0208

[88] G. I. Zekos. 2006. State Cyberspace Jurisdiction and Personal Cyberspace Jurisdiction. *International Journal of Law and Information Technology* 15, 1, 1–37. http://doi.org/10.1093/ijlit/eai029

[89] Shanyang Zhao, Sherri Grasmuck, and Jason Martin. 2008. Identity construction on Facebook: Digital Empowerment in Anchored Relationships. *Computers in Human Behavior* 24, 5, 1816–1836. http://doi.org/10.1016/j.chb.2008.02.012

Threats to Autonomy from Emerging ICTs

Brandt Owen Dainow
Maynooth University
Ireland
bd@thinkmetrics.com

Abstract

This paper examines threats to autonomy created by significant emerging ICT's. Emerging ICT's cover a wide range of technologies, from intelligent environments to neuroelectronics, yet human autonomy is potentially threatened by all of them in some way. However, there is no single agreed definition of autonomy. This paper therefore considers the ways in which different versions of autonomy are impacted by different systems. From this range of threats we will derive some properties which any ICT must exhibit in order to threaten human autonomy. Finally, we will show how the range of definitions of autonomy creates problems for customary approaches to vale-sensitive design, and indicates a need for greater flexibility when attempting to improve the ethical status of emerging ICT's.

Keywords: ICT; ethics; autonomy; foresight studies; futures research; ETICA; value-sensitive design

1 Introduction

This paper examines ways in which human autonomy can be threatened by emerging ICT's. Emerging ICT's embrace a wide range of technologies, including autonomous systems, intelligent environments, bio-electronic implants, robotics, and artificial intelligence. We will explore the different ways in which each ICT threatens autonomy to show that the nature of the threat, if any, depends on the version of autonomy being used. From this survey of threats we will derive some properties which any ICT must exhibit if it threatens human autonomy, no matter how it is defined. Finally, we will argue the range of definitions of autonomy demonstrates the need for greater flexibility when attempting to improve the ethical status of emerging ICT's.

In order to discuss the complete range of emerging ICT's it is necessary to have a system by which to conceptualise and organise them. We shall use for these purposes the output of the ETICA project (Stahl 2011), which attempted an examination of the complete range of emerging ICT's, developed an ordered taxonomy (Ikonen et al. 2010), and researched the ethical concerns associated with each (Heersmink, van den Hoven, and Timmermans 2010).

2 Methodologies for evaluation of emergent technology

To count as "emerging" (as opposed to "possible") an ICT must be sufficiently developed that we can understand its general features, but which has not yet achieved maturity in all aspects. Since the 1960's the predominant model for such analysis has thus been to treat technologies as "socio-technical systems" (Trist 1981; Lamb and Kling 2003). This position contrasts with the other major treatment of technology, technological determinism, which treats technology as affecting society, but as developing independent of social, cultural, political and other "intangible" causes (Smith and Marx 1994). An *emerging* ICT will have little market penetration (it may still be in prototype), such that its final place and role in the market is yet to become evident; its full and final set of features are unlikely to be completely determined; the business models under which it will function are likely to be nascent; users may still be working out patterns of usage; and the regulatory framework is most probably undeveloped.

Autonomous cars provide a simple example of an emerging ICT which illustrate these uncertainties. Google's autonomous cars are still in prototype (Solveforx 2016), while Tesla vehicles are just beginning to penetrate the market (Statistica 2016). In both cases, the systems are still in development such that the final set of functions is indeterminate (Solveforx 2016; Musk 2016). The regulatory framework for autonomous vehicles has seen some development in a few places, such as California, but most countries haven't even begun to think about the technology. In the USA, for example, as of March 2017, only eleven states had any regulations regarding this technology, while the majority of attempts to introduce legislation in the USA since 2012 have failed (National Conference of State Legislatures 2017). Business models for the sale and use of this technology are still undetermined. For example, Tesla recently proposed models in which people would be able to send their Tesla car out as an automated taxi when they're not using it (Musk 2016). In the meantime, users are still learning how to use the systems, often in conflict with how the designers intend them to be used. The most dramatic example of this disparity between designer's intent and user behaviour is the frequent driving of Tesla vehicles without human supervision, despite the Tesla's frequent statements that the user should pay attention and keep their hands on the steering wheel at all times (Consumer Reports 2016). Any attempt to understand the ethical impact of autonomous cars must therefore commence with a vision of the state of affairs when the technology has matured – with standardised features, customary patterns of usage, stable regulatory regimes and (reasonably) settled business models. Any attempt to assess the ethical implications of autonomous cars thus becomes an exercise in anticipating what the future will look like. This brings us into the remit of foresight studies.

2.1 Introducing Foresight Studies

'Foresight studies' is one of the terms used to describe the discipline of considering the future in a methodical manner. Foresight studies can be used in any discipline and so the objectives in such research vary widely (Glenn 2009; Godet 2009; Veikko, Kanerva, and Kouri 2009). This presence within many disciplines has led to a variety of terms being used to describe the activity. Common alternative terms for foresight studies are "future studies" (most commonly used in the USA), "futures research" (Europe), "futurology" (Australasia), "prospective studies" (France), "futures field" (Europe) and "prognostics" (Russia) (Veikko, Kanerva, and Kouri 2009). A wide variety of foresight methodologies have evolved and most foresight research projects combine several methodologies.

The different methodologies used within foresight studies can be divided between quantitative and qualitative in technique and between normative and exploratory in purpose, though some methodologies bridge these divides (Veikko, Kanerva, and Kouri 2009; Haegeman et al. 2013; Gordon and Glenn 2004). Quantitative methods are used mainly in predictive forecasting and seek to produce probabilities (Veikko, Kanerva, and Kouri 2009; Haegeman et al. 2013). Quantitative methods are also used in complex modelling, for example, climate modelling. Such methods are based on the assumption that future development is a continuation of past trends. While they provide a continuity from past to future, they cannot allow for disruptive technologies or unexpected developments. Some argue for the need to support quantitative methods with qualitative ones (Veikko, Kanerva, and Kouri 2009; Haegeman et al. 2013) as they are better suited for the anticipation of abrupt or unexpected changes (Ogilvy 2002; Veikko, Kanerva, and Kouri 2009). Foresight research also distinguishes between methodologies on the basis of whether they are normative or exploratory. Normative forecasting is concerned with whether a particular future is desirable and the ethical status of decisions which would lead to it. By contrast, exploratory forecasting explores what is possible, irrespective of its desirability (Veikko, Kanerva, and Kouri 2009). Ethical assessment of the future is rarely a formal component of foresight research project (Bell 1996; Poli 2011; Brey 2012; Clarke 2005; Grunwald 1999; Harris et al. 2010). Key foresight research figures such as Wendell Bell have attempted to determine what ethical values are universally accepted by all peoples in all cultures (Bell 1996). However, Bell has since been criticised for accepting as objective and universal what are merely western cultural values (Poli 2011; Dator 2011; Rubin 2011). A less controversial approach has been to incorporate ethical values as research data. Here distinct sets of exploratory and normative data are gathered. For example, the functionality of future technologies can be drawn the visions of those creating them, after which philosophical sources can be mined for ethical treatments of these functions. This was the approach taken by the ETICA project.

2.2 ETICA

The most comprehensive foresight research project focused on ethical assessment of emerging ICT's so far has been the EU's ETICA (**Et**hical Issues of Emerging **IC**T **A**pplications) project, which ran from 2009 - 2011. The project's aims were to identify "significant emerging ICT's" (Stahl 2011, 1), the ethical issues these gave rise to, and to make recommendations for EU policy makers. A key output from the ETICA project was a taxonomy of emerging ICT's. Here ETICA developed formal structures for technological descriptions and scenario construction (Ikonen et al. 2010). It then used bibliometric analysis to assess published discourse regarding the ethical issues associated with these ICT's (Heersmink, van den Hoven, and Timmermans 2010). The project validated these findings with expert focus groups and surveys (Heersmink et al. 2010). Finally, ETICA developed a series of recommendations (Rainey and Goujon 2011), together with supporting philosophical and methodological expositions (Veikko, Kanerva, and Kouri 2009; Rader et al. 2010; Rainey and Goujon 2011).

ETICA identified 107 different technologies as significant and emerging, which it aggregated into eleven groups. It then identified approximately 400 ethical concerns associated with these technology groups. The aim was to collate the ethical concerns, not explore them in depth, much as an 18th century naval cartographer might map the location and outlines of a chain of islands, but not explore their interior. It is not our intention to review all of these ethical concerns, but to concentrate on issues concerning autonomy. We shall first catalogue how human autonomy can be challenged by each technology group, then elucidate the common characteristics shared by all. However, the meaning of the term 'autonomy' is much contested, so what one considers a threat to autonomy depends on what one means by the term. It is therefore necessary to briefly explore the concept of autonomy. Our aim is not to produce a definition of the term 'autonomy', but to better understand the different ways it may be used when discussing ethical threats.

3 Autonomy

Emmanuel Kant was the first to apply the concept of autonomy to human beings (Schneewind 2007; Dryden 2015; Paul, Miller, and Paul 2003; Dworkin 1988). Before Kant the term 'autonomy' was a purely political one. A state was said to be "autonomous" if its laws were drafted within that state, as opposed to being drafted by a distant imperial court or some similar extra-national body (Dworkin 2015; Schneewind 2007). Kant applied the concept to the individual in *Groundwork of the Metaphysics of Morals* (Kant 1785). His aim was to justify the right of each individual to make their own judgements regarding morality. For the previous 100 years, Enlightenment thinkers had been seeking arguments to combat the politically dominant understanding of ethics, in which ordinary people were seen as too weak-willed to act morally without threats of punishment and promises of reward. Under this view, society was dependant on the guidance of those few exceptional people whom God had enabled to understand and teach His moral laws. The essence of morality for everyone else was to do what they were told (Schneewind 2007). *Groundwork of the Metaphysics of Morals* set out to prove that each person drafted their own rules for their own conduct, that these rules constituted each person's moral code, and that no education was required for this ability - it was universally present in all humans. Kant's labelled the innate capacity of all humans to determine what was morally correct as 'autonomy.' He then used this capacity for autonomy as the basis for human dignity, which then became the basis for human rights. Thus Kant's concept of autonomy created an interlinking of the concepts of individuality, freedom, morality, autonomy, dignity and politics (Schneewind 2007; Bittner 2014; Dryden 2015; Paul, Miller, and Paul 2003).

Since Kant introduced the concept, a number of differing definitions of autonomy have arisen. The range of such accounts, as well as the lack of progress towards any consensus, has led to the concept of "regimes" of autonomy (Anderson 2014; Dworkin 2015) – clusters of similar, but not identical, theories, such that "about the only features held constant from one author to another are that autonomy is a feature of persons and that it is a desirable quality to have." (Dworkin 2015, 8). Kant's criteria for autonomy were, by later standards, rigid and limited. Under Kant's account, people were required to determine what was right or wrong without reference to the consequences of their actions. Furthermore, under Kant, rules of conduct only counted as a moral if the person believed everyone should act the same way. There was no room for moral pluralism in Kant, or tolerance of other people acting differently (Kant 1998; Guyer 2003; Johnson 2014). This type of definition of autonomy has come to be known as a "substantive procedural" account. A "procedural" account of autonomy takes the position that autonomy is a specific form of thinking process (ie: follows a specific procedure). A subset of procedural accounts are substantive, adding the requirement that autonomy also involves thinking about certain things. By contrast, "content-neutral" procedural accounts try to define autonomy without reference to any particular concern or aim. Content-neutral accounts of autonomy form the majority of modern accounts (Dryden 2015) and are, perhaps, more suited to multi-cultural or pluralistic societies in that they allow individuals to determine for themselves what sources and values to consider when making ethical decisions.

The value given to autonomy varies according to the importance given to competing claims, most frequently those of paternalism and those forms of communitarianism which devalue individual choice in favour of community needs (Bell 2014). Kant's original concept was developed in an effort to remove external authority as a source of morality. Kant specifically identifies autonomy as oppositional to outside influence. However, we are all influenced by external forces, including family, cultural and religious influences. Consequently, outside influence and the point at which it reduces autonomy is a problem which must be dealt with by every definition of autonomy. One response has been to argue that the concept of autonomy as self-governance isolates the individual from their society and morally devalues family, community, culture and tradition (Christman 2014; Mackenzie and Stoljar 2000; Donchin 2000; Buss and Zalta 2015). These positions have given rise to relational and social conceptions of autonomy based on the necessity of human sociality.

It has also been argued autonomy, if conceived of as only certain forms of thinking, does not account for the full range of human emotional and physical experience, such as trained muscle action (e.g.: playing sport or music) or healthy reflexes, which can also express personal autonomy (Meyers 2005, 1989). These issues collide with particular urgency where ICT's become concerned with care for the ill or aged (Agich 2003; Burmeister 2016; Dworkin 1988), but also wherever ICT's come to mediate what someone considers to be core elements of their life.

Kant applied autonomy strictly to the moral realm, though it had political implications. Since then Western Philosophy has come to term the capacity to determine one's own moral codes as 'moral autonomy.' To moral autonomy have been added the concepts of "personal autonomy" (the capacity to determine one's own actions) and "political autonomy" (the capacity to make one's own political decisions and have them heeded) (Dryden 2015; Anderson and Christman 2005). What these three definitions share is the concept of "self-governance." ICT's threaten human autonomy whenever they interfere with this self-governance. This occurs whenever ICT's make decisions for people, especially when an ICT imposes on the user a way of doing something which is antithetical to the manner in which the user would have chosen. Some threats to autonomy are universal under all definitions of autonomy, but most depend on the particular form of autonomy being used. Reducing autonomy may be justified on other grounds but such a justification, like the threat, will be based on the version of autonomy being used. It will often be the case that a justification works under one concept of autonomy, but not under others.

4 ETICA's technology groups and their threats to autonomy

Identification of ethical issues of significant emerging ICT's is dependent upon first having identified the ICT's which are both significant and emerging. ETICA defined "significant" technologies as those which were likely to have an important influence on society and conceptions of the self. It defined "emerging" technologies as those on a stable development path and likely to be in common use by 2030 (Stahl 2011). 107 technologies meeting these criteria were identified, including 3-D printing, sensor networks, augmented reality, location-based services, lab-on-a-chip, molecular electronics, nanobots, mobile payment systems, robotics, social network analysis, speech recognition, internet of things, and wireless power. These were then organised into eleven groups on the basis of shared characteristics, such as type of user, operational context, functionality, and relation to other technologies. For example, both wireless sensor networks and wearable computing were placed into the category of Ambient Intelligence.

The project identified the following eleven technology groups:

1. Affective computing
2. Ambient intelligence
3. Artificial intelligence
4. Bioelectronics
5. Cloud computing
6. Future internet
7. Human/machine symbiosis
8. Neuroelectronics
9. Quantum computing
10. Robotics
11. Virtual and augmented reality

(Ikonen et al. 2010, 42; Stahl 2011, 5)

Not all technology groups were seen by ETICA as generating ethical threats in and of themselves. ETICA assessed quantum computing as unlikely to have any impact on society in the next 10-20 years, while cloud computing and future internet were seen as enablers of other technologies and not as raising unique ethical issues themselves. We shall now briefly examine the remaining technology groups and discuss some of the ways in which each has the potential to reduce human autonomy. Our aim is not to explore the details of the individual threats, but simply to see how autonomy can be threatened by each technology. While a few technologies threaten autonomy no matter how it is defined, the central point of this survey is to demonstrate that most threats only appear under specific definitions of autonomy.

4.1 Affective Computing

Affective computing involves treating human emotion as input and generating output which either emulates human signals of emotion or which seeks to manipulate human emotions. Here ethical concerns derive from the importance of human emotion in communication and the risks of manipulation and misunderstanding.

How affective computing threatens autonomy depends on whether affective computing works or not. Unlike some other technologies, there is no guarantee it will ever be possible to build systems which can accurately read human emotions (Barrett, Gendron, and Huang 2009; Picard 2003). The greatest danger with affective computing lies in the possibility that people will think it is working when it is not. We can usually tell when most technologies operate incorrectly. However, there is no simple way of checking to see if an affective computing system's assessment of a person's emotion is accurate. Incorrect assessment of someone's emotions threaten their autonomy when that data is used to make decisions which affect them.

If ICT's can successfully understand human emotions, affective computing threatens privacy through the ability to glean information about people which they do not wish revealed. Wherever such information is used to make decisions affecting the person, it may constitute a restriction of their autonomy. Here it depends on the form of autonomy one subscribes to. Some "second-order" procedural accounts of autonomy hold that autonomy is preserved provided one would have agreed with the decision if they had been given the chance (Dworkin 1988, 2015). However, accounts of autonomy which do not allow for such "second-order" assessments hold that any decision denied the person is a restriction on their autonomy. If those decisions were made in an effort to benefit the person, they *may* be justifiable on the grounds of form of paternalism. Otherwise they simply constitute a form of oppression.

Affective computing therefore constitutes a threat to autonomy whenever affective data is used to make decisions about someone, unless one is using a second-order procedural definition of autonomy *and* the person would have agreed with the decision if they had been given the chance.

Emulation of emotion makes possible threats to autonomy in that it offers the chance to overwhelm rationality with emotion. Any manipulation of a person's emotions which prompts them to make a decision they would not have made otherwise constitutes a restriction of autonomy, irrespective of how autonomy is conceived. The concern is therefore that introducing affective output changes the power balance between system and user. If the ability to accurately detect emotions is combined with the ability to emulate emotional expression, the power of such systems to manipulate and persuade becomes greater than any previous human invention. In this sense, affective systems become human manipulation technology.

4.2 Ambient Intelligence

Ambient intelligence refers to IC technology which is embedded into the environment. Its defining characteristic is the invisibility of the devices, often associated with automated input (for example, a change in room temperature) and less-than obvious responses by the ambient system (for example, changing the air conditioning settings).

Ambient intelligence offers the capability of personalising the environment to the individual user. Where personalisation is under control of the user, it represents an extension of their autonomy. Where personalisation is outside the user's control, it represents a reduction in autonomy under most, but not all, definitions. Second-order accounts allow for the preservation of autonomy if the user would have agreed to the personalisation had they been given the chance. Only under such accounts is autonomy preserved by automated environmental personalisation. If the personalisation is intended to improve the user's quality of life, reductions in autonomy may be justifiable on the grounds of paternalism. However, that is a problematic debate which cannot be resolved in terms of universal principles, but will ride on the specifics of each case; the type of personalisation, the individuals involved and their personal values.

Ambient intelligence shares several ethical issues with affective computing. Autonomy is threatened by poor quality personalization services, either as a result of an inadequate understanding of human nature, or by the skewing of personalisation protocols resulting from commercial interest or developer ignorance (or bias). Inference of user needs from their behaviour is particularly problematic, especially with regard to children. There is the danger that users will be forced to change their behaviour to get the best out of ambient systems, and so end up being "trained" by their environment's designers (Soraker and Brey 2007).

Ambient intelligence creates the risk of the personal environment acquiring power over people. Since ambient intelligence makes the personal environment controllable by third parties, the personal environment becomes a contestable zone, open to commercial and state interests, both in terms of gathering information and seeking control. Some contestation of the personal environment has always occurred. For example, disputes over noise can be traced back to the earliest cities (Goldsmith 2012). However, ambient intelligence represents such a significant increase in power that contestation for the personal environment becomes a new type of issue – the power to (potentially) totally control someone else's personal space.

4.3 Artificial Intelligence

ETICA examined both "hard" AI (emulation of human thought) and "soft AI" (e.g.: expert systems and software agents). Ethical concerns focused on the use of soft AI as an enabler for other emerging technologies, such as ambient intelligence and affective computing. While issues associated with hard AI were considered, ETICA's judged that hard AI was unlikely to do more than appear in earliest prototypes by 2030 and so was outside the scope of the project.

Our discussions of ambient intelligence and affective computing have both shown how human autonomy can be threatened by the power of these systems over people. AI significantly enhances this power, and so can be an enabler of threats to autonomy wherever it is deployed within ICT systems. Particularly problematic is the possibility that AI may enable systems to learn things about people not anticipated by designer or user. Implementation of informed consent becomes extremely difficult when one does not know what information may be acquired. It has been suggested the solution is to build ethical reasoning into AI systems (Dennis et al. 2013). However, ethical systems are abstract principles of value, not problem-solving algorithms (Kraut 2016; Graham 2004). Ethical dilemmas only exist where multiple ethical values conflict because it is these conflicting values themselves which create the dilemma. AI developers are not, therefore, able to reach for an existing corpus of ethical problem-solving algorithms as if the matter were a simple sort task. Any implementation of ethical processing will therefore impose someone's particular ethical values on the system. Should such an ethical value set not accord with the user's set of ethical values, the user's autonomy will be compromised by the actions of their AI-enabled system. The only way to counter such a threat would be to allow users to customise their AI-enabled systems with their own ethical values and reasoning.

4.4 Bioelectronics

ETICA defined bioelectronics as ICT systems which interact directly with the human body. The primary ethical considerations were focused on use in health care, but those who see bioelectronics as a path to artificial human bodies were also considered. Here bioelectronics represents the potential for technological modification of the human form. It therefore directly confronts the essence of what it is to be human and impacts core human values, such as autonomy, freedom and dignity.

Bioelectronics can be used to change people's internal states by delivering medicines and intervening in bodily processes and thus has the potential to reduce autonomy. However, this concern is dependent on the concept of autonomy used. If one accepts a second-order procedural concept of autonomy (Dworkin 1988), autonomy may be preserved *post-hoc* by agreeing with the earlier bioelectronic intervention. However, conceptions which situate autonomy at the moment of decision must see any bioelectronic intervention as a reduction of autonomy. Whether this loss of autonomy can be justified on paternalistic grounds depends on the intent behind the intervention. However, it may be possible to reframe the issue such that autonomy is not threatened. Discussions of bioelectronics tend to treat intervention as morally equivalent to an intentional act. Under this perspective, each individual bioelectronic intervention is treated as a discrete act to which assent may be given or withheld. This raises problems for the preservation of autonomy because of the requirement for ongoing consent (Agich 2003). However, where this intervention is frequent and automated, such as is the case with drug delivery, it may be more productive to think of it as a form of reflex. Some definitions of autonomy incorporate reflex into their schema as non-cognitive expressions of the self (Meyers 2005, 1989). Treating bioelectronic interventions as artificial reflexes removes any threat to autonomy under such definitions.

The nature of bioelectronic intervention means it is difficult to resist. This places a great deal of potential power in the hands of those controlling the devices; granting them the ability to control someone to a degree not possible by other means. The capability of bioelectronics to change internal bodily states and processes is the same as that of drugs. Accordingly, bioelectronics shares many of the same issues pertaining to autonomy as do drugs. There are today debates about the advisability of treating some mental states as illnesses, concerns about inappropriate use of drugs to control behaviour in elderly and children, and other issues relating to appropriate boundaries to medical intervention (Smith 2012). These same issues apply to interventions by bioelectronic devices. Threats to autonomy apply here in two ways; firstly, "authentic" conceptions of autonomy involve the concept of authenticity (that there is some given essential nature to each individual), such that autonomy requires being in conformance to that essential self (Kühler and Jelinek 2013). These accounts hold that there is something innate about humans which cannot be changed without harming their authenticity, such as particular forms of cognition (Frankfurt 1971). Theological accounts of autonomy may also depend upon a divinely-ordained authentic human nature (Wilson 1978; Niebuhr 2004). Meanwhile, "coherentist" accounts base autonomy on continuity with personal history, such that changes or decisions which radically depart from someone's previous patterns or tastes are inauthentic, and thus reductions in autonomy (Miller 1981; Ekstrom 1993). Bioelectronic interventions which take a person away from an essential self, however that self is conceived, are reductions to these forms of autonomy. Finally, where bioelectronic interventions are unwelcome and forced on the individual, they constitute a clear reduction in autonomy under all definitions of autonomy.

4.5 Human-Machine Symbiosis

ETICA characterised human-machine symbiosis as the pairing of innate human capabilities with ICT. This definition covers an extremely wide range of systems, including haptic interfaces, decision-support systems, computer-assisted surgery, augmented reality and direct neural interfaces. There is considerable overlap with other technologies, especially bio- and neuroelectronics and artificial intelligence.

Attempts to enhance human beings through the addition of ICT components may constitute reductions in autonomy under the authenticist views just outlined. Therapeutic devices, designed to replace lost human capabilities rather than enhance them, may be autonomy-preserving, but other accounts can hold that even these constitute a loss of autonomy (Bublitz and Merkel 2009). This is particularly the case where those devices require monitoring or control by others. A device may therefore reduce autonomy when it is operated by someone else, but enhance autonomy where it is controlled by the user (Sharon 2017). Such concerns are not limited to devices implanted into people's bodies. Use of external devices, for example, as aids to memory, may constitute reductions to autonomy, especially under accounts based on the concept of an "extended self" (Olson 2011; Rachlin and Jones 2010), in which parts of our personal identity are embedded in external objects, such as clothing or mobile phones (Ahuvia 2005; Belk 1988; Turkle 2011). For someone who embeds part of their identity in their phone, even something as simple as an enforced software update may constitute a reduction in their autonomy.

4.6 Neuroelectronics

ETICA defined neuroelectronics as technology interfacing between the human nervous system and electronic devices. While neuroelectronics clearly overlaps with bioelectronics, ETICA felt the intimate connection between the person and their brain meant neuroelectronics could not be adequately examined within a general treatment of bioelectronics.

As with bioelectronics, autonomy is not threatened when neuroelectronics is used to gather data, but only when used to induce changes. Since changes cannot be made without knowledge of internal states, gathering data is, however, a privacy concern as an enabler of reductions to autonomy. Any neuroelectronic system which produces changes within the person constitutes a threat under most accounts of autonomy. While this is clearly the case with conceptions of autonomy which involve an authentic nature or coherent life-history, it may also be the case under second-order procedural accounts (which allow for autonomy preservation if someone agrees to a change afterwards). This is because a person's agreement to a previous procedure may be merely the result of the changes they have undergone. Where these devices are installed to deal with cognitive impairment, such as dementia, informed consent is impossible and thus autonomy must be reduced (though this may be justifiable on paternalistic grounds). As with bioelectronics and human-machine symbiosis, any accounts of autonomy which involve a divinely-derived authentic human nature will view all neuroelectronic changes as reductions in autonomy.

4.7 Robotics

ETICA defined robots as "machines with motor function that are able to perceive their environment and operate autonomously" (Ikonen et al. 2010, 114). ETICA's focused on new developments which increase the mobility and intelligence of robotic devices, permitting their deployment into wider areas of society, such as the home and healthcare, and focused on robots built for specific roles within these contexts. ETICA did not consider general-purpose, fully mobile devices controlled by strong artificial intelligence seeking to make war on humanity.

The degree to which robotics can threaten autonomy depends on the definition of autonomy used and the use made of the robot. As has been noted earlier, traditional procedural accounts of autonomy as self-determination have been criticised for unrealistically portraying people as isolated individuals and leaving no space in the account for communitarian elements such as the influence of family or culture (Mackenzie and Stoljar 2000; Stoljar 2015). On this basis some have gone so far as to argue autonomy is an unattainable ideal (Strawson 1994). A more common response has been to develop a treatment of autonomy which includes space for the influence of others, such as the concept of a "social self" (Meyers 2005, 44). The central premise in relational accounts of autonomy is the necessary role of other people in the development of one's values and the centrality of interpersonal relationships to human existence; that it is not possible to function autonomously without the influence of others (Mackenzie and Stoljar 2000; Stoljar 2015; Donchin 2000). Such conceptions of autonomy are threatened when robots replace humans in roles which have social externalities, such as health care and education. While robots are able to undertake the central tasks, the loss of the human contact constitutes a reduction in relational autonomy. The consequence of this is that there may be some tasks for which robots are not ethically suitable, for no other reason than that they are robots. If the social externalities of a work role are essential for the maintenance of the recipient's autonomy, then that role must be reserved for humans. Were we to replace nurses or teachers with robots, we could thus see people argue they have an inalienable human right to refuse to be served by a robot and to demand human service.

4.8 Virtual/Augmented Reality

ETICA's union of virtual reality with augmented reality into one technology group was based on the common feature they share - the imposition of digital output onto the human sensory field. ETICA defined virtual reality as occurring where digital output completely replaced sensory data. Where it did not completely replace external sensory data, ETICA used the term 'augmented reality' (Heersmink, van den Hoven, and Timmermans 2010, 114). ETICA's concerns were founded on the fact that such systems mediate or replace interaction with the physical environment.

Autonomy is threatened when virtual or augmented realities depict objects, people and places in the real world and thus become involved with the many ethical issues associated with depiction, cultural bias and its influence the development of attitudes and taste (Zimbardo and Leippe 1991). The imposition of foreign cultural models in virtual realities would constitute a threat to relational and social conceptions of autonomy (Tomlinson 1991), while lack of alternatives and lack of configuration options threatens all conceptions of autonomy.

4.9 Summary

Many of the technologies examined provide personalised services tailored to the individual. These technologies threaten human autonomy when they fail to personalise effectively. This may result from misreading the user, insufficient granularity within the personalisation, lack of user control, inadequate modelling of the user or by imposing on the user ways of living which are contrary to their values. A number of the technologies change power balances within society, granting for the first time, or greatly increasing, the actant power of the environment over the individual. This power threatens autonomy if used to impose on the user or to enable autonomy-limiting personalisation. Some conceptions of autonomy contain elements which can be disrupted by some ICT's. Conceptions of autonomy which include inter-personal relationships are challenged by robotics, while conceptions of autonomy which make reference to authenticity, coherent life-history, or some form of given human nature, are challenged by bio- and neuro-electronics and may also be challenged by human-machine symbiosis.

Irrespective of the nature of autonomy used, our survey of ETICA's research suggests there are certain characteristics any ICT is likely to have if it threatens autonomy:

1. **Surveillance**: Surveillance consists of the obtaining of information pertaining to an individual from their behaviour or communication, followed by the use of that data for purposes, either unknown to that individual or against their wishes (Lyon, Ball, and Haggerty 2012). While it can be used in a paternalistic fashion to enhance the individual's autonomy, our concern is when it is used in a manner which limits the individual's autonomy. The presence of a surveillance system is a necessary precondition for the use of personal information by others and for personalisation services because it constitutes the means by which information supporting personalisation is gathered. Control of the outflowing data by the individual is therefore an effective way in which to retain autonomy. Thus control of privacy is an enabling right to autonomy.

2. **Disparity of Control**: Disparity of control occurs when a third party has a greater control over the use of one's personal data and the autonomy-limiting processes than oneself. This lack of control can be accomplished through lack of knowledge on the part of the individual as to what is occurring, through "take it or leave it" terms of use combined with a lack of alternatives, and through the lack (or concealment) of user configuration capabilities. In some cases, such as data profiling, not just functions, but entire industries may be concealed (Federal Trade Commission 2014; Turow 2011; Dainow 2015).

3. **Insufficient Configurability**: Autonomy can be threatened by lack of variability in configuration options suitable to reflect the variations in the user's desires, style of operation or range of outcomes. This lack of variation can stem from:

   - A lack of recognition by the ICT developers of the need for, or even existence of, such variations.
   - Insufficient consideration of need for user variability during the design process.
   - Insufficient granularity of configuration options.
   - Systemic biases within the delivery mechanisms, including business models, market competition, regulatory framework or any other factors regarding the operational delivery of services. For example, a system might be highly configurable, but delivered to users with a standardised configuration which cannot be changed in an effort to reduce support costs.

4. **Insufficient variation in operational models**: Many new ICT's are owned and operated by a very limited number of providers (Noam et al. 2003; Hillis, Jarrett, and Petit 2013). Often a single provider dominates the market to near-monopoly levels. This limits choice of service provision, and consequently the model under which it is provided. In many cases the business model is identical across all providers within an ICT sector, such that change of provider does not change the circumstances under which the user accesses the service. For example, social networking is only available under a capitalist for-profit model in which service provision is exchanged for personal data which is then commoditised (or monetised). Non-profit and privacy-preserving social network systems do not exist to the degree that a real choice is available. In addition, IP protection promotes walled gardens and suppresses interoperability, further limiting options for choice by users because it is not possible to interoperate between social network providers.

5 Conclusions

The common feature to all conceptions of autonomy is self-determination. What one self-decides, how, on what basis, according to what procedure, and subject to which influences, varies from account to account. Yet efforts to preserve autonomy are central to ICT ethics (Spiekermann 2015). Efforts to reduce the negative ethical impact of ICT's on people through processes such as value-sensitive design need to be cognisant of the various conceptions of autonomy relevant to the intended functionality of the system. This is especially the case with technologies which enter into spheres of life which have previously been purely human or which reduce the power people have over their personal lifeworld. Given the wide range of definitions of autonomy, no single definition can possibly cover all cases.

Because autonomy is frequently central to the ethical status of any technology, it is rarely possible to categorise any ICT as threatening or reducing autonomy without reference to the form of autonomy used. Methodologies to improve the ethical sensitivity of ICT technology, such as value-sensitive design, make frequent reference to the need to specify a particular brand of ethics, such as deontology or utilitarianism, from which to draw values (Spiekermann 2015; Nissenbaum 2001; Winfield, Blum, and Liu 2014), but the application of such values still depends on the concept of autonomy being used. It is generally assumed that there is a single, fixed meaning to autonomy, at least in any given context, and that preservation of autonomy simply requires, at most, identification of the correct version (Manders-Huits 2011). Some have noted the meaning of autonomy can vary with different contexts (Burmeister 2016), but it is possible that it only appears the meaning has changed because the difficulties perceived apply to one version of autonomy and not others, thus making the problematic version of autonomy obvious. Our review of ETICA's survey of threats to autonomy has revealed that many versions of autonomy may be applicable within the one context, some of which may have problems while others do not. Thus, attempts to find the "correct" version of autonomy for any context may be moot. On this basis, it seems more appropriate to develop ICT functionality which does not depend on a single definition of autonomy, but adapts to the user's own definition. This implies a loose pairing of system behaviour to coding, such that output can be fine-tuned after deployment, either by a user, their technical support staff or through the use of soft AI components focused on adapting the system to user feedback.

All significant emerging ICTs can threaten autonomy, but there is no simple way of avoiding this. Efforts such as value-sensitive design stand more chance of success if they focus less on incorporating a single set of values and focus instead on ways by which users may adapt ICT's to reflect their own individual values.

References

Agich, George J. 2003. *Dependence and Autonomy in Old Age: An Ethical Framework for Long-Term Care*. 2nd ed., Rev. Cambridge, UK ; New York: Cambridge University Press.

Ahuvia, Aaron C. 2005. 'Beyond the Extended Self: Loved Objects and Consumers' Identity Narratives'. *Journal of Consumer Research* 32 (1): 171–184.

Anderson, Joel. 2014. 'Regimes of Autonomy'. *Ethical Theory and Moral Practice* 17 (3): 355–68. doi:10.1007/s10677-013-9448-x.

Anderson, Joel, and John Christman. 2005. 'Introduction'. In *Autonomy and the Challenges of Liberalism: New Essays*, 1–23. Cambridge, UK ; New York: Cambridge University Press.

Barrett, Lisa Feldman, Maria Gendron, and Yang-Ming Huang. 2009. 'Do Discrete Emotions Exist?' *Philosophical Psychology* 22 (4): 427–37. doi:10.1080/09515080903153634.

Belk, Russell W. 1988. 'Possessions and the Extended Self'. *Journal of Consumer Research* 15 (2): 139–168.

Bell, D. 2014. 'Communitarianism'. In *The Stanford Encyclopaedia of Philosophy*, edited by Edward N. Zalta, Summer 2014. http://plato.stanford.edu/archives/sum2014/entries/communitarianism/.

Bell, Wendell. 1996. *Foundations of Future Studies*. New Jersey: Transaction Publishers.

Bittner, Rüdiger. 2014. 'Autonomy Modest'. *Erkenntnis* 79 (7): 1329–39.

Brey, Philip A. E. 2012. 'Anticipatory Ethics for Emerging Technologies'. *NanoEthics* 6 (1): 1–13. doi:10.1007/s11569-012-0141-7.

Bublitz, Jan Christoph, and Reinhard Merkel. 2009. 'Autonomy and Authenticity of Enhanced Personality Traits'. *Bioethics* 23 (6): 360–374.

Burmeister, Oliver K. 2016. 'The Development of Assistive Dementia Technology That Accounts for the Values of Those Affected by Its Use'. *Ethics and Information Technology*, 1–14. doi:10.1007/s10676-016-9404-2.

Buss, Sarah, and Edward N. Zalta. 2015. 'Personal Autonomy'. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/personal-autonomy/.

Christman, John. 2014. 'Relational Autonomy and the Social Dynamics of Paternalism'. *Ethical Theory and Moral Practice* 17 (3): 369–82. doi:10.1007/s10677-013-9449-9.

Clarke, Steve. 2005. 'Future Technologies, Dystopic Futures and the Precautionary Principle'. *Ethics and Information Technology* 7 (3): 121–126.

Consumer Reports. 2016. 'Tesla's Autopilot: Too Much Autonomy Too Soon'. Consumer Reports. http://www.consumerreports.org/tesla/tesla-autopilot-too-much-autonomy-too-soon/.

Dainow, B. 2015. 'Digital Alienation as the Foundation of Online Privacy Concerns'. *Computers & Society* ETHICOMP Special Issue: 109–17. doi:10.1145/2874239.2874255.

Dator, Jim. 2011. 'Wendell Bell: The Futurist Who Would Put My Grandmother in Prison'. *Futures* 43 (6): 578–582.

Dennis, Louise, Michael Fisher, Marija Slavkovik, and Matt Webster. 2013. 'Ethical Choice in Unforeseen Circumstances'. In *Towards Autonomous Robotic Systems*, 433–445. Springer.

Donchin, Anne. 2000. 'Autonomy and Interdependence: Quandaries in Genetic Decision-Making'. In *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, edited by Catriona Mackenzie and Natalie Stoljar, 236–58. New York: Oxford University Press.

Dryden, Jane. 2015. 'Autonomy'. *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/autonomy/.

Dworkin, Gerald. 1988. *The Theory and Practice of Autonomy*. Cambridge Studies in Philosophy. Cambridge ; New York: Cambridge University Press.

———. 2015. 'The Nature of Autonomy'. *Nordic Journal of Studies in Educational Policy* 1 (0). doi:10.3402/nstep.v1.28479.

Ekstrom, Laura Waddell. 1993. 'A Coherence Theory of Autonomy'. *Philosophy and Phenomenological Research* 53 (3): 599–616.

Federal Trade Commission. 2014. 'Data Brokers: A Call for Transparency'. Federal Trade Commission. https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

Frankfurt, Harry G. 1971. 'Freedom of the Will and the Concept of a Person'. *The Journal of Philosophy* 68 (1): 5–20.

Glenn, Jerome C. 2009. 'Introduction to Futures Research Methods'. In *Futures Research Methods*, edited by Jerome C Glenn and Theodore J Gordon, 3rd ed. Washington, DC: The Millenium Project.

Godet, Michel. 2009. '11. Structural Analysis'. In *Futures Research Methods*, 3rd ed. Washington, DC: The Millenium Project.

Goldsmith, Mike. 2012. *Discord: The Story of Noise*. Oxford: Oxford University Press.

Gordon, Theodore J, and Jerome C Glenn. 2004. "Integration, Comparisons, and Frontiers of Futures Research Methods'. In *EU-US Seminar: New Technology Foresight, Forecasting & Assessment Methods*. Vol. 1314.

Graham, Gordon. 2004. *Eight Theories of Ethics*. London; New York: Routledge/Taylor and Francis Group.

Grunwald, Armin. 1999. 'Technology Assessment or Ethics of Technology? Reflections on Technology Development between Social Sciences and Philosophy.' *Ethical Perspect* 6 (2): 171 – 182.

Guyer, Paul. 2003. 'Kant on the Theory and Practice of Autonomy'. In *Autonomy*, edited by Ellen Frankel Paul, Fred Dycus Miller, and Jeffrey Paul, 70–98. Cambridge: Cambridge University Press.

Haegeman, Karel, Elisabetta Marinelli, Fabiana Scapolo, Andrea Ricci, and Alexander Sokolov. 2013. 'Quantitative and Qualitative Approaches in Future-Oriented Technology Analysis (FTA): From Combination to Integration?' *Technological Forecasting and Social Change* 80 (3): 386–397.

Harris, Ian, Penny Duquenoy, Richard Jennings, David Pullinger, and Simon Rogerson. 2010. 'Ethical Assessment of New Technologies: A Meta-Methodology'. *Journal of Information, Communication and Ethics in Society* 9 (1): 49–64.

Heersmink, Richard, Jeroen van den Hoven, and Job Timmermans. 2010. 'D.2.2 Normative Issues Report'. D.2.2. ETICA Project.

Heersmink, Richard, Job Timmermans, Jeroen van den Hoven, and Kutoma Wakunuma. 2010. 'D.2.3 Normative Issues Workshop Report'. D.2.3. ETICA Project.

Hillis, Ken, Kylie Jarrett, and Michael Petit. 2013. *Google and the Culture of Search*. New York: Routledge.

Ikonen, Veikko, Minni Kanerva, Panu Kouri, Bernd Stahl, and Kutoma Wakunuma. 2010. 'D.1.2. Emerging Technologies Report'. D.1.2. ETICA Project.

Johnson, Robert. 2014. 'Kant's Moral Philosophy'. Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/sum2014/entries/kant-moral/.

Kant, Immanuel. 1785. *Grundlegung Zur Metaphysik Der Sitten*. Riga: Johann Friedrich Harttnoch.

———. 1998. *Groundwork of the Metaphysics of Morals*. Translated by Mary J. Gregor. Cambridge Texts in the History of Philosophy. New York: Cambridge University Press.

Kraut, Richard. 2016. 'Aristotle's Ethics'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2016. http://plato.stanford.edu/archives/spr2016/entries/aristotle-ethics/.

Kühler, Michael, and Nadja Jelinek, eds. 2013. *Autonomy and the Self.* Dordrecht: Springer Netherlands. http://link.springer.com/10.1007/978-94-007-4789-0.

Lamb, Roberta, and Rob Kling. 2003. 'Reconceptualizing Users as Social Actors in Information Systems Research'. *MIS Quarterly*, 197–236.

Lyon, David, Kirstie Ball, and Kevin D Haggerty. 2012. *Routledge Handbook of Surveillance Studies*. Abingdon, Oxon; New York, NY: Routledge.

Mackenzie, Catriona, and Natalie Stoljar, eds. 2000. *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. New York: Oxford University Press.

Manders-Huits, Noëmi. 2011. 'What Values in Design? The Challenge of Incorporating Moral Values into Design'. *Science and Engineering Ethics* 17 (2): 271–287. doi:10.1007/s11948-010-9198-2.

Meyers, Diana T. 1989. *Self, Society, and Personal Choice*. New York: Columbia University Press.

———. 2005. 'Decentralizing Autonomy: Five Faces of Selfhood'. In *Autonomy and the Challenges to Liberalism*, edited by John Christman and Joel Anderson, 27–55. Cambridge: Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9780511610325A011.

Miller, Bruce L. 1981. 'Autonomy & the Refusal of Lifesaving Treatment'. *Hastings Center Report* 11 (4): 22–28.

Musk, Elon. 2016. 'Tesla Master Plan, Part Deux'. *Personal Blog*. July 20. https://www.tesla.com/blog/master-plan-part-deux.

National Conference of State Legislatures. 2017. 'Autonomous | Self-Driving Vehicles Legislation'. National Conference of State Legislatures. http://www.ncsl.org/research/transportation/autonomous-vehicles-legislation.aspx.

Niebuhr, R. 2004. *The Nature and Destiny of Man: A Christian Interpretation : Human Nature*. Library of Theological Ethics. Westminster, John Knox Press.

Nissenbaum, H. 2001. 'How Computer Systems Embody Values'. *Computer* 34 (3): 118–20. doi:10.1109/2.910905.

Noam, Eli M, Donald Hay, Michael R Baye, and John Morgan. 2003. 'The Internet: Still Wide Open and Competitive?' *OII Internet Issue Brief*, no. 1.

Ogilvy, James. 2002. 'Futures Studies and the Human Sciences: The Case for Normative Scenarios'. In *New Thinking for a New Millennium: The Knowledge Base of Futures Studies*, edited by Richard Slaughter. London: Routledge.

Olson, Eric T. 2011. 'The Extended Self'. *Minds and Machines* 21 (4): 481–495.

Paul, Ellen Frankel, Fred Dycus Miller, and Jeffrey Paul, eds. 2003. *Autonomy*. Cambridge: Cambridge University Press.

Picard, Rosalind W. 2003. 'Affective Computing: Challenges'. *International Journal of Human-Computer Studies* 59 (1): 55–64.

Poli, Roberto. 2011. 'Ethics and Future Studies'. *Int. J. Management Concepts and Philosophy* 5 (4): 403–10.

Rachlin, Howard, and Bryan A. Jones. 2010. 'The Extended Self.' In *Impulsivity: The Behavioral and Neurological Science of Discounting.*, edited by Gregory J. Madden and Warren K. Bickel, 411–31. Washington: American Psychological Association. http://content.apa.org/books/12069-015.

Rader, Michael, A. Antener, Rafael Capurro, Michael Nagenborg, and L. Stengel. 2010. 'D.3.2 Evaluation Report'. ETICA Project.

Rainey, Stephen, and Philippe Goujon. 2011. 'D.4.2 Governance Recommendations'. D.4.2. ETICA Project. http://www.academia.edu/download/30848560/Deliverable4.2.pdf.

Rubin, Anita. 2011. 'Living in the Age of Emotional Rationality: Wendell Bell, Social Media and the Challenges of Value Change'. *Futures* 43 (6): 583–589.

Schneewind, Jerome B. 2007. *The Invention of Autonomy: A History of Modern Moral Philosophy*. 7th ed. Cambridge: Cambridge Univ. Press.

Sharon, Tamar. 2017. 'Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare'. *Philosophy & Technology* 30 (1): 93–121. doi:10.1007/s13347-016-0215-5.

Smith, Brendan L. 2012. 'Inappropriate Prescribing'. *Monitor on Psychology* 43 (6): 36.

Smith, Merritt Roe, and Leo Marx. 1994. *Does Technology Drive History?: The Dilemma of Technological Determinism*. Cambridge, Mass.: MIT Press.

Solveforx. 2016. 'Google Self-Driving Car Project Website'. *Google Self-Driving Car Project*. https://www.google.com/selfdrivingcar.

Soraker, Johnny Hartz, and Philip Brey. 2007. 'Ambient Intelligence and Problems with Inferring Desires from Behaviour'. *International Review of Information Ethics* 8: 7–12.

Spiekermann, Sarah. 2015. *Ethical IT Innovation: A Value-Based System Design Approach*. Boca Raton, Fl.: Taylor & Francis.

Stahl, Bernd Carsten. 2011. 'Project Final Report'. ETICA Project.

Statistica. 2016. 'Statistics and Facts about Tesla'. *The Statistics Portal*. http://www.statista.com/topics/2086/tesla/.

Stoljar, Natalie. 2015. 'Feminist Perspectives on Autonomy'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2015. http://plato.stanford.edu/archives/fall2015/entries/feminism-autonomy/.

Strawson, Galen. 1994. 'The Impossibility of Moral Responsibility'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 75 (1/2): 5–24.

Tomlinson, John. 1991. *Cultural Imperialism: A Critical Introduction*. Baltimore, Md.: Johns Hopkins University Press.

Trist, Eric L. 1981. *On Socio-Technical Systems*. Ontario: Ontario Ministry of Labour.

Turkle, Sherry. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books. http://public.eblib.com/choice/publicfullrecord.aspx?p=684281.

Turow, Joseph. 2011. *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your World*. New Haven: Yale University Press.

Veikko, Ikonen, Minni Kanerva, and Panu Kouri. 2009. 'D.1.1 Heuristics & Methodology Report - Final'. D.1.1. ETICA Project.

Wilson, Edward O. 1978. *On Human Nature*. Cambridge, Mass.: Harvard University Press.

Winfield, Alan FT, Christian Blum, and Wenguo Liu. 2014. 'Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection'. In *Advances in Autonomous Robotics Systems*, 85–96. Springer.

Zimbardo, Philip G, and Michael R Leippe. 1991. *The Psychology of Attitude Change and Social Influence*. Philadelphia: Temple University Press.

ORBIT

Smart City Transcendent

Understanding the smart city by transcending ontology

Dainow, Brandt

National University of Ireland, Department of Computer Science.

Corresponding Author: Brandt Dainow, brandt.dainow@nuim.ie

Abstract

This paper provides a conception of the smart city which takes into account what the smart city brings into the world which is new and original. This approach provides a means of dealing with the complex influences humans and digital systems will have on each other in the mature smart cities of the future. I will first review traditional accounts of the smart city and derive from them the essential characteristics common to these visions. I will then show how these characteristics can be best understood through Actor-network theory and construct an account of the smart city as an autopoietic system in which humans and devices are co-constituting actants. Finally, I shall develop this into an original conception of the smart city as a new type of thing - an "integrated domain."

Keywords: smart city, computer, society, actor-network theory, autopoiesis, future internet

Accounts of Smart Cities

There is no single definition of what constitutes a smart city. Accounts divide into three schools regarding definitions of the smart city. One school defines the smart city in terms of what we can do with it. Here we find topics such as innovation policy (Komninos, Schaffers, & Pallot, 2011), smart governance (Vinod Kumar, 2015), urban planning (Zygiaris, 2013), improved sustainability (Bowerman, Braverman, Taylor, Todosow, & Von Wimmersperg, 2000) and similar large-scale management tasks. Others define the smart city in terms of its material construction. Here the focus is mainly on devices comprising the Internet of Things and their interactions. Some focus on issues of designing smart sensors and other components, such as weight detectors in roads (Hancke, Silva, & Hancke, 2013). Others focus on the interactions between components, such as sensor networks (Filipponi et al., 2010) and human-sensor interactions (Pettersson et al., 2011). Others attempt to make sense of both usage and components through the development of organisational models. For example, Jin proposes a four-layer model of sensors, networks, cloud-based data management and service delivery (Jin, Gubbi, Marusic, & Palaniswami, 2014). Organising smart city components and operation into layers is fairly common. Balakrishna's four-layer model (Balakrishna, 2012) is almost identical to Jin's, while others adopt a three-layer (Atzori, Iera, Morabito, & Nitti, 2012) or five-layer (Pettersson et al., 2011) model. In contrast, some have attempted more conceptual organisational models. For example, Filiponi (Filipponi et al.,

2010) suggests a vertical model based on our understanding of the space in which activity occurs, such as car, home, office and civic space.

Attempting to make sense of this range, *Smart Cities: Definitions, Dimensions, Performance, and Initiatives* (Albino, Berardi, & Dangelico, 2015) examined twenty-two accounts and cross-referenced these with urban developments which labelled themselves as smart cities. Their conclusion was that a single definition was impossible because of wide variations in the meaning of common terms. Attempts to include everything inevitably cover too much or too little. At one extreme we have very broad statements, such as a smart city is that which uses "social, mobile and sensor-based technologies … to create more productive alignments between%1 (growing) demand and (constrained) resources." (Hartswood, Grimpe, Jirotka, & Anderson, 2014, p. 3). At the other extreme, *Getting Smarter About Smart Cities* defines a smart city as:

"using networked, digital technologies and urban big data to tackle a range of issues, such as improving governance and service delivery, creating more resilient critical infrastructure, growing the local economy, becoming more sustainable, producing better mobility, gaining transparency and accountability, enhancing quality of life, and increasing safety and security." (Kitchin, 2016, p. 9).

Deployment of smart city technology is more likely to be accomplished by industry than by academic researchers. It is therefore worth considering the vision of the smart city which industry offers. Cobham Plc is a major player in the smart city sector, with annual turnover of $US3 billion (Cobham PLC, 2011). Their Tactical Communications and Surveillance division provides a range of smart city technologies. Their smart city sales brochure (Cobham PLC, 2014) offers a 5-layer model, ranging from an "IP mesh" which allows any type of sensor to communicate with any other, through several layers of device type, including integration of personal devices as sensors (with and without user knowledge), through to management systems.

Certain characteristics are held in common throughout all these accounts. They agree that smart cities require a ubiquitous heterogeneous sensor network which provides information about the inhabitants, their environment and service delivery. This has been referred to as an "IP mesh" (Cobham PLC, 2011, p. 3) and as an "underlying sensor fabric" (Balakrishna, 2012, p. 224). These sensors must report their data to other devices. Some of these communication patterns can be predesigned, while others must be created on an ad hoc basis in response to the movement of the inhabitants and machines such as drones, robots and cars (Guo, Wang, Zhang, Yu, & Zhou, 2013; Pettersson et al., 2011). There is general agreement these devices will be embedded in the civic environment, the home and other personal spaces (such as the car), worn on the person and implanted within the body. While many accounts assume that all data processing will occur in the cloud, this is not inevitable. A contrasting view can be seen in the concepts of fog computing (Bonomi, Milito, Zhu, & Addepalli, 2012; Petrolo, Loscrì, & Mitton, 2015) and user-controlled personal data stores (Service Systems Group, 2015b). Under these views significant data processing can be done locally, either within sensor devices themselves or through locally situated data processing units (Service Systems Group, 2015a). Indeed, such local processing is likely to be more secure and efficient (Dainow, 2015; Langheinrich, 2001).

The aim of this paper to provide a comprehensive definition of the smart city which can serve as the foundation for ethical analysis by reconceptualising the smart city. Current ethical concerns for the smart city tend to be limited to specific issues associated with specific technologies within the smart city, such as cars (Jaisingh, El-Khatib, & Akalu, 2016) or location detection (Martínez-Ballesté, Pérez-Martínez, & Solanas, 2013). My aim is to provide a comprehensive framework for analysis of any ethical issue, especially issues which are emergent from the interaction of multiple systems and which therefore cannot be predicted from any one sub-system. This requires developing a new vision of the smart city which is able to account for the complex nature of interactions within it.

The City of the Future

We must first place the smart city within a future studies framework. This is because the smart city does not yet exist, but is rather in an early stage of development. Given that the smart city does not yet exist, current technologies related to smart cities must be regarded as interim steps towards what will be the ultimate deployed solutions in mature smart cities of the future. We must therefore regard current smart city technologies as in flux, as unfinished, and their features as mutable and almost certainly subject to change. Furthermore, our current understanding of the path to the mature smart city is highly uncertain. The issue of ad hoc networking illustrates that even the direction of innovation is in dispute. Much concern in smart city development focuses on issues of communication between devices. It is widely recognised that mobile devices, autonomous systems and moving people will all necessitate the dynamic creation and uncreation of unpredictable network patterns. Known as "ad hoc networking", some see great advantages in this (Boldrini, Conti, Delmastro, & Passarella, 2010; Guo, Wang, Zhang, Yu, & Zhou, 2013) seeking only to make it reliable and secure, while others (Pettersson et al., 2011) see its chaotic and unpredictable nature as something which needs to be controlled. Clearly, it is difficult to anticipate the ultimate role of ad hoc networking in the smart city at a stage when we cannot even agree on whether it should be promoted or inhibited.

Teleological accounts, concerned with urban governance, typically focus on the ways in which smart cities can solve today's problems. However, it is unlikely that the contentious issues of smart city governance are predictable in any detail. Prior to the invention of the world wide web, no one could have anticipated the need for domain name registries, nor the political fighting which would emerge around them. We can therefore anticipate that at least some important forms of governance of mature smart cities are yet to be imagined. Similarly, the most pressing issues confronting mature smart cities will include those we cannot yet anticipate because they will derive from the unknown nature of currently unpredicted technologies and the currently unpredictable patterns of usage which will evolve around them. Considering that new technologies always create negative side-effects (Cardwell, 1994; Derry & Williams, 1993), it is inevitable that the solutions we deploy in creating the smart city will generate problems of their own – which we also cannot predict.

Placing the smart city within a futures perspective therefore means understanding that we do not know how it will be made, how it will operate, or how it will be managed. Conceptions of this future unknown smart city must therefore frame themselves in terms which are not dependent upon knowledge of the specific details of future smart technologies. Fundamental to the futures approach is an understanding that the smart city is not today's city with some digital tech laid over the top of it, any more than the modern city is just a Victorian one with cars instead of horses. The mature smart city will be a fundamentally different type of environment from any we have previously seen. The key change will be the presence of ubiquitous ICT technologies. Through all of history the human built environment has been a largely dumb. It has not had the ability to do things to us except in the most gross fashion and it has not had the capability to know us. The capacity of the environment to obtain data and to respond to human action represents a fundamentally new type of built environment for human living.

We can therefore understand the characteristics of the smart city best if we consider it as a form of built environment. Accounts of smart cities focused on the technical infrastructure typically treat it as consisting of intelligent devices embedded within dumb materials. However, the number, ubiquity, heterogeneity and invisibility of most devices will cause humans to understand and act towards the object in which the devices are embedded rather than the devices themselves, and in many cases deal with aggregations of devices rather than individual ones. Attempting to account for this interaction between human and device in terms of component types and processes is impossible on two grounds. Firstly, we cannot know the nature of these future devices or what patterns of usage humans will develop towards them. Secondly, the number and heterogeneity of these devices and variability of human relations with them across differing contexts renders attempts to understand on the basis of technical type impossibly complex. We must therefore cease to talk in terms of technical components and instead identify the functional characteristics which can unite the various heterogeneous technical forms.

The following will identify the essential functional characteristics of the smart city. Thereafter I will use these to synthesise a conception of the smart city as a new type of being.

Ambient Intelligence

'Ambient intelligence' is defined as technology which embeds input, processing or response ubiquitously through the environment (Ikonen, Kanerva, Kouri, Stahl, & Wakunuma, 2010). With its ubiquitous sensor mesh the smart city is the ideal type for ambient intelligence. As a lived experience, people will know the smart city not as individual components and discrete processes isolated within atomised sections of their lives. Instead the smart city will be experienced as a seamless experience as one moves from house to car to work, from one context to another. People will not think of themselves as moving from one discrete situation to another, they will simply experience themselves as going about their daily life, an unbroken stream of changing contexts seamlessly merging from one to the other. The following quotes from the auto industry illustrate current steps in this direction:

Matt Jones, Director of Future Technology at Jaguar Land Rover:

"We very quickly discovered that customer expectations really aren't based on our Jaguar Land Rover competition - they're based on the smartphone. They're based on the tablet. They're based on the home entertainment experience…. Consumers expect to be able to download an app, get into the car and discover a seamless integration between the device and the automobile." (Bedigian, 2016, p. 7)

John Schnoes, Programme Director of Vehicle Information Technology at Nissan:

"It all comes down to smart device connectivity. People are really looking to have that sort of seamless experience when they've been working on their phone and then they walk into their car and they expect the platform to know what they've been doing." (Bedigian, 2016, p. 8)

Emergent in-car systems reveal that cars will communicate with their environment in a deep fashion. Usage-Based Insurance systems analyse driver behaviour, including time of day, acceleration rates, cornering and locations travelled in order to dynamically set insurance premiums. Emerging collision-detection systems can automatically contact emergency services with essential details, including the identities of the occupants and which seatbelts were fastened. Car maintenance systems will access user calendars to determine the best time for maintenance appointments. Already today, the Ford SyNC systems can accesses the driver's medical data to remind them when to take their medicine (Jaisingh et al., 2016).

We can see similar cross-context connectivity trends at the urban management level. For example, Restore NV provides demand management systems, such as those controlling electrical grids, including five of Europe's largest grid operators (Restore NV, 2016). Their newly announced smart city electricity management system, FlexPond, monitors individual home appliances to predict electrical load within individual homes (Restore NV, 2017). This information can be cross-referenced with local weather patterns and the number and type of household occupants to greatly improve demand prediction (Hancke et al., 2013).

Other smart city systems treat personal digital devices as mobile components of the smart city infrastructure. This is not limited to using such devices merely as sensors or as personal identifiers. Patterns and content of social interaction between people can be analysed to anticipate movement patterns, shared use of resources, and even as a way of determining appropriate security levels to create between devices (Atzori et al., 2012). Others have tested analysis of social networking to guide structuring of ad hoc networks (Boldrini et al., 2010). China plans to take this further, creating a "social credit system" by monitoring many aspects of personal conduct, including honesty and conformance to socialist behaviour. A key aim is to reduce social diversity. This system will offer rewards and punishments through differential access to smart city services (State Council of People's Republic of China, 2014). Some smart city services will even penetrate the physical body. Already being tested, body sensor networks embedded into the environment communicate with implanted medical devices, such as heart monitors, to anticipate and react to medical emergencies (Lo, Thiemjarus, King, & Yang, 2005). Such health sensor networks can also monitor some vital signs externally, such as skin temperature, respiration, sleep patterns and diet (Hancke et al., 2013; Y. Kim, Kim, & Lee, 2008).

The picture which emerges is one in which smart city services must take data from a ubiquitous digital ecosystem, in which digital devices are embedded throughout the fabric of the built environment - "transforming everyday objects into information appliances," (Botta, de Donato, Persico, & Pescapé, 2016, p. 691) but also including devices carried by people and embedded within their bodies (Balakrishna, 2012; Botta et al., 2016; Filipponi et al., 2010; Jin et al., 2014; Pettersson et al., 2011). This data will be integrated across contexts, technology forms and purposes so as to create an integrated sensing and response environment (Psyllidis, 2015). Such an "intelligent information infrastructure" (ITU-T, 2008, p. 2) is clearly an ideal type for ambient intelligence.

Artificial Intelligence

Central to the vision of the smart city is algorithmic intelligence, or "soft AI" (e.g.: expert systems and software agents) (Ikonen et al., 2010; Komninos, 2006). It has been estimated that a typical smart city will contain around 1 trillion nanoscale devices (Balakrishna, 2012). A central paradigm implied within the concept of the smart city is that the smart city will generate new data and novel possibilities for action as a result of the integration of data from disparate domains (Picon, 2015). There is a general acceptance that this will require machine learning and other forms of soft AI (Balakrishna, 2012; Nam & Pardo, 2011). In addition to common expectations that soft AI will provide core functionality through cloud computing and big data, it has been suggested soft AI will need to be embedded at local level to support ad hoc networking and to facilitate more rapid system responses due to the sheer scale of data (Komninos et al., 2011).

Robotics

Robotics refers to ICT devices possessing the ability to move autonomously (Ikonen et al., 2010). Self-driving cars (and possibly drones) are expected to exist within smart cities (Balakrishna, 2012; Jaisingh et al., 2016; Petrolo et al., 2015). The home will see robotic devices such as vacuum cleaners and similar devices, sometimes known as mobile IoT (Gubbi, Buyya, Marusic, & Palaniswami, 2013). The specific details of which robotic devices will exist within the smart city need not concern us for the purposes of this analysis. The important point is that the smart city environment will involve both humans and machines moving within the same spaces. Just as the introduction of the car eventually led to the rise of pedestrian crossings, so an environment of moving machines will require adjustments in human behaviour to take them into account.

The smart city as a socio-technical system

The vision which has emerged of the smart city places technical artefacts and humans as equally powerful actants. Digital devices will communicate amongst themselves and engage in negotiations (Atzori et al., 2012; State Council of People's Republic of China, 2014) as they create ad hoc networks, including event-driven networks (Filipponi et al., 2010), contextual networks (Boldrini et al., 2010) and social networks (Atzori et al., 2012; Boldrini et al., 2010). Thus the network structure of the smart city will be created by the devices as well as by the humans. Device activity will respond to human activity. Humans will respond to these responses. Furthermore, digital devices will have their own needs, which will require consideration by humans. Accordingly we can view digital devices as causal agents in a smart city on the same causal level as humans.

Systems theory has dominated urban planning since the 1960's (Taylor, 2005) and permeates approaches to the smart city (Söderström, Paasche, & Klauser, 2014), which is seen as a complex system, or system of systems (Albino et al., 2015; Chourabi et al., 2012). Actor-network theory (ANT) (Latour, 2005) provides a theoretical framework for incorporating the smart city's digital agents and their needs into an interactive relationship with humans. By treating both devices and humans as co-existing within the same structure, ANT provides a framework by which we can attribute to artefacts causative properties within the human dimension and vice versa (Tabak, 2015). A further advantage of ANT is the lack of need to specify the details of each node within the network (Law, 1992). This suits a both a heterogeneous device mix and a futures approach which accepts that we cannot know what final forms smart city technology will take. In addition, ANT has been used to account for pathways in ICT innovation (R. Kim & Kaplan, 2005; Lamb & Kling, 2003) and so applying it to smart cities can be linked to that research. This offers the potential for an explanatory framework which can account for the dynamics of the innovation which will lead to the mature smart city.

The arrival of an intelligent responsive environment, especially one containing autonomously mobile agents, requires changes which must affect people. This will inevitably bring the needs of the digital agent into conflict with the needs of the individual or society, just as cars require roads and regulations which gives them effective rights over people in some circumstances. Ambient intelligence systems may also require changes in human behaviour, raising the possibility that our environment will train us to suit its needs (Soraker & Brey, 2007). In other cases, we can anticipate that personalisation of shared spaces will result in differences between people regarding how much any particular personalisation suits them. This may range from the trivial, such as ambient temperature, to the existential, such as access to sensor networks required for medical implants. As the Chinese social credit system shows, some people will be forced to make compromises in order to accommodate aspects of personalisation preferred by others within the shared space. It is not inevitable that spaces will be personalised to suit the majority, or that minority needs will be accommodated. It is possible that environmental personalisation will become a zone of contention and power dynamics. Given the potential impact alteration of spaces can have on the individual, control of personalisation of shared spaces could easily become a path to domination of others. Hence control of the digital actants within the smart city's actor-network can be expected to grant influence and power over the human actants.

The fact digital actants have their own needs which require changes by humans raises the issue of secondary rights. Digital objects, while requiring humans change, do not, in and of themselves, originate that need. They are operational units whose existence is caused by humans and to whom the benefit of their activity is given to humans. They are not self-originating and they are not in receipt of the benefit of their activity. As a result, while it appears that the object is competing with the person, in actual fact the object stands in proxy to another individual - the beneficial owner. While the situation may look like competition of rights between the individual and the object, it is in fact competition between one individual and another in which the device stands as proxy for the second individual. In such circumstances it is not a new situation. However, the rise of AI systems may mean that, while other humans benefit from the satisfaction of digital needs, they did not originate those needs. Instead we may encounter needs that were generated by the AI system as a result of its own development and self-learning. Given the expected complexity and scale of a smart city's ambient intelligence, there are likely to be many contexts in which we cannot distinguish between human-originated and self-determined needs of digital actants. It is likely that some needs will arise via a combination of human-originated and self-determined needs. It is likely that many human goals will be accomplished via methodologies which were self-determined by autonomous systems. Where a self-determined methodology has negative effects on other people, we can expect some form of resistance. Such a situation represents an interactive process between human and digital system, in which human decision-making contends with the autonomous decisions of a digital actant. In that ANT considers nodes as "black boxes" (Tabak, 2015, p. 37) and does not need to specify their internal details, it is not confounded by the issue of whether an activity derives from the device or some beneficial owner.

By treating the human and the digital components of the smart city as equal actants within the same environment, ANT also offers a dimension of analysis which is essential for a full understanding of the interactions of the human and the digital – the perspective of the digital device. By its very nature, a digital device cannot know the physical world. What a device knows is what its sensors generated as input. For example, a thermostat does not respond to the temperature but what the sensors tell it the temperature is. If the sensor malfunctions, or we game it, the thermostat control will still respond to what the sensor tells it, not what the material reality is. Hence digital devices do not know the physical world, but inhabit a totally digital environment. What they know of us and how much they know is determined by our representation in digital terms. Smart city systems will not respond to us; they will respond to our digital representations. If those digital representations are incorrect or incomplete, then the analysis of us will be compromised and the delivery of services will suffer.

Accordingly, ANT offers us a conception of the smart city as a complex network of heterogeneous actants. Under this view we recognise that these actants may be a single digital device, an individual person, a group of people, a group of devices, or a mixture of both. We can build on this conception by considering the logic which must be inherent within operational smart cities in order for them to exist. Doing so will provide a more detailed understanding of the nature of the smart city system, revealing its essential quality of autopoiesis.

Autopoiesis in the Smart City

The concept of autopoiesis originates in theoretical biology in the 1970's (Varela, Maturana, & Uribe, 1974) and accounts for living things, such as cells and bodies, as self-sustaining systems. The concept was made accessible to the social sciences by Luhmann (Luhmann, 1986) in the 1980's. Systems theory holds that the defining characteristics of a system are not determined by the properties of the components but by the patterns of their relationships (Varela et al., 1974). Autopoietic systems are those which are self-maintaining in the face of changing stimuli, either changes inside the system or within the system's environment. To be autopoietic, systems must also regenerate the essential patterns which characterise that system and internally generate the components and processes required to maintain it (Maturana, 1981).

The original theories of autopoiesis were designed for biology and seemed limited to that field by the necessary characteristic that an autopoietic system generate its own components (Maturana, 1981; Varela, Maturana, & Uribe, 1981), a process known as "material self-production" (Di Paolo, 2005, p. 433). However, this issue is really one of the minimum granularity in one's ontological schema. Living systems are the ideal type for theories of autopoiesis, but they do not create atoms or many of the molecules they depend on. No autopoietic system generates the materials from which the components considered necessary for characterisation of that system as autopoietic are created. All autopoietic systems exchange input and output with their environment (Luhmann, 1986). Hence autopoietic systems need only self-generate the components which are the most proximate cause of their characteristics and self-maintenance.

Building on this approach, Luhmann reworked the concept of autopoiesis to bring it into the social sciences (Luhmann, 1986). Under his account social systems are autopoietic systems which use communication as their characteristic form of autopoietic generation. The atoms from which social systems are constituted are communications, which are recursively produced and reproduced. Luhmann defines communications as being composed of three elements; information, utterance and understanding. As such, a communication is an event. A social system as autopoietic is a "network of events which produces itself … the reproduction of events by events." (Luhmann, 1986, p. 175). The critical point is that the elements of an autopoietic system which make it autopoietic are those which maintain autopoiesis in the face of destabilising forces and which give rise to characteristics of autopoiesis which distinguish it from other autopoietic systems. The basic unit of social autopoiesis is this communicative triad of information, understanding, and utterance (Luhmann, 1986).

This pattern is easily visible within the smart city. It is the minimum necessary characteristic of digital device interaction able to constitute any of the essential operational characteristics which can constitute a defining characteristic of an ambient intelligence. Within ambient intelligence this communicative pattern is visible as the ability to accept input ("information"), process it in some way ("understanding") and consequently produce output ("utterance"). In order for any device to participate in an ambient intelligence it must be able to perform these three processes. The actant within the autopoietic system is thus defined by its ability to accept input, process it and respond. The response becomes input for a connected node and thus the process perpetuates.

However, smart cities will not be constituted by closed communicative systems in which digital devices communicate only with each other. As we have seen, the smart city will be characterized by the close integration of digital devices and humans (Bicocchi et al., 2013). Much of the digital device's input will derive from humans, both as intentional commands and as unconscious input (such as movement and reactions to digital service delivery). Hence both devices and humans will react to each other and stimulate new responses in turn. In this way both humans and digital devices will be seen to engage in communicative patterns amongst their own kind and across the human-digital divide. Furthermore, once begun, this system becomes self-referential and recursive, as each responds to the response of the other, and so the system becomes self-sustaining and autopoietic.

If we accept this communicative triad as the atomic unit of a living smart city, there exists a temptation to see this as two societies, one human and one digital, communicating within themselves and only interacting with the other by means of constituting its environment – humans embedded in a digital environment and digital devices surrounded by humans. However, significant portions of human communicative action will be mediated by the digital environment, while that digital environment's communicative patterns will respond in myriad ways to every human action. Under such circumstance, few communicative triads can be said to be purely human or digital; the overall patterns of communicative flow will be the result of both human and digital components. Some have highlighted the tight interaction of the ICT and human communities in smart cities under the terms 'collective sensing', 'collective awareness' and 'collective action' (Bicocchi et al., 2013). However, such analysis does not incorporate the deep fusing of both collectives which must occur. I am not suggesting this fusion constitutes some form of hybrid human-machine society, or that we should redefine the term 'society' to include digital systems. Some have used the term 'smart society' to characterise the society within a smart city (Hartswood et al., 2014), but this refers to a human society using smart city services. This is a useful definition to maintain because it refers to real issues and to a clearly definable zone of concern. I therefore believe we need a new term for this new phenomena, one which refers to an autopoietic system constituted of communicative triads between nodes which may be digital, human or both. I propose to call this an 'integrated domain'.

Defining integrated domains

An integrated domain is a socio-technical system in which humans and digital devices co-mingle in a manner such that it becomes impossible (or even meaningless) to identify the origins of patterns within the system as being either human or digital. An integrated domain is autopoietic. The autopoiesis of an integrated domain derives from a close coupling of nodes such that output from one node cannot avoid generating a reaction in another node, combined with the unavoidable structure of the atomic node, which is the communicative triad event. An integrated domain consists of two collectives integrated into a mutually dependant and co-creative partnership. One collective consists of a human smart society. This is a form of human society existing within an ambient digital environment, such that human perceptions, actions and intersubjectivity are unavoidably mediated and influenced by this ambient digital environment. The other (non-human) collective consists of an autopoietic system of digital devices and networks based on the communicative triad. However, neither collective possesses strict boundaries against the other, but rather the two intermingle. We can thus describe each as separate for theoretical purposes, but it is not possible to account for characteristics of, or phenomena within, either without reference to nodes within the other. Any discussion of the nature and processes within smart society on the human side or the smart city digital components must draw on aspects of the other, and so any explanatory discussion of either must occur at the higher ontological level of the integrated domain.

The concept of autopoiesis includes a persistent demand for regeneration (Varela et al., 1981). This holds true for integrated domains as well. However, whereas in a living organism it is the material constituents which need regeneration, in an integrated domain it is the communicative triads. Communicative triads are events, not states (Luhmann, 1986). Patterns of communication thus form an actor-network in which the nodes are events constituted of communicative triads, whose material base may be a human, a digital device, or a group of either or both. It is essential that groups of people and/or devices can be treated as equal actants within this model because groups will interact and communicate with individual devices or people in the smart city, and vice versa. An example of a combined group which would yet constitute a single node is a car in motion. The car is composed of a wide range of interacting devices. Much of the operation of these devices will be influenced by the actions of the driver, and vice versa. However, this complex system may communicate data aggregated from many of these systems in combination with the driver's actions to a single device, such as a traffic light. Similarly, a single device, such as a traffic light, can be expected to communicate with dozens, if not hundreds, of similarly complex entities. We thus see that the communicative system is integrated not just across differences of materiality, but also spans (or ignores) multiple ontological levels.

Social machines

The primary unit of the integrated domain is the communicative triad. This is a three-stage process consisting of input, processing and output. It constitutes an individual node within the network structure of the integrated domain. This network structure is dynamic, variable and self-sustaining. Under this view the smart city is seen as a system possessing autopoiesis. Each node is not a state, but a process or an event and their influence on the system's patterns is not determined by their material base. The material base of each node may be an individual processing unit, such as an single digital device or an individual person, or it may be constituted by a grouping of devices, people or even a combination of both. Each node may also be constituted by the aggregated activity of a group of sub-nodes. These sub-nodes may themselves may be constituted by another group, an overlapping set of groups or an individual person or device. It is likely that the cascade effect of events and responses, combined with the volume and complexity of communication will make it impossible, or even meaningless, to ask whether a feature of note is the product of one node or many, human or digital. The communicative triad therefore constitutes both the atomic unit for, and the essential pattern of, interactions within the smart city. It provides us with a framework for handling the way in which smart city processes can transcend normally distinct ontological boundaries and frequently renders them meaningless. This deep interconnectedness of materially indeterminate nodes combines with ad hoc networking and human response variability to create a highly complex network structure.

As an autopoietic system, many of the communicative triads will be concerned with internal states of operation within both digital assets and humans. In the case of digital assets, much of the communication can be expected to be concerned with issues such as maintenance and management of digital technology. This monitoring represents self-awareness. This is not awareness in the human sense, implying sentience. However, as a system, the integrated domain is aware of its internal processes. Through the component of the human, the integrated domain possesses the human level of understanding and meaning-giving. An integrated domain is therefore a self-aware, autopoietic system composed of actants who may be digital or human, individual or group.

We are already witnessing early examples of hybrid systems which combine human and digital devices under the label of “social machines”. Social machines are defined as “socio-technical systems which involve the participation of human individuals and technological components…able to extend the reach of both human and machine intelligence [by] supporting capabilities that less integrated systems might find difficult to accomplish” (Smart, Simperl, & Shadbolt, 2014, pp. 55–56). Social machines are the early fore-runners of the integrated domain which will form the fabric of future smart cities.

Conclusions

The close integration of a huge number range of devices and systems makes any model of the smart city which depends upon material differences impossibly complex and inevitably incomplete. The deep integration of various ontological levels, from the individual nano-sensor to the global cloud, makes it impossible to comprehend causes of individual actions within the smart city. Both of these issues are further magnified by the expectation and need of the human lived experience to transcend these technical distinctions. Discussion of smart cities in terms of enablers of human activity, such as urban management, do not adequately incorporate into their models the deep fusion of digital cognitive processes with human deliberations. This confronts us with the necessity for a model which does away with such distinctions. Instead we have postulated the smart city as an “integrated domain”. Under this view the digital and the human, the individual and the group, all hold equal status as nodes within a complex, dynamic autopoietic system known as a smart city. The concept of the smart city as an integrated domain provides a framework through which it is possible to consider issues which transcend traditional IT boundaries, issues deriving from complex interactions of multiple agents of multiple types. It provides a unified model by which to account for bidirectional interactions between management-level civic aims and individual device functions, between groups and individuals, between digital devices and humans. It further provides us with the ability to anticipate issues within mature smart cities without knowing the details of the technologies to come. Models like this, which seek to incorporate the human and the digital into integrated systems, offer our best hope of anticipating the future issues of smart cities.

References

Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart Cities: Definitions, Dimensions, Performance, and Initiatives. *Journal of Urban Technology*, *22*(1), 3–21. https://doi.org/10.1080/10630732.2014.942092

Atzori, L., Iera, A., Morabito, G., & Nitti, M. (2012). The social internet of things (siot)–when social networks meet the internet of things: Concept, architecture and network characterization. *Computer Networks*, *56*(16), 3594–3608.

Balakrishna, C. (2012). Enabling Technologies for Smart City Services and Applications. In *2012 Sixth International Conference on Next Generation Mobile Applications, Services and Technologies* (pp. 223–227). https://doi.org/10.1109/NGMAST.2012.51

Bedigian, L. (2016). *TU Automotive Detroit 2016 Conference Report*. Detroit: TU-Automotive Ltd (Penton).

Bicocchi, N., Cecaj, A., Fontana, D., Mamei, M., Sassi, A., & Zambonelli, F. (2013). Collective awareness for human-ict collaboration in smart cities. In *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2013 IEEE 22nd International Workshop on* (pp. 3–8). IEEE.

Boldrini, C., Conti, M., Delmastro, F., & Passarella, A. (2010). Context-and social-aware middleware for opportunistic networks. *Journal of Network and Computer Applications*, *33*(5), 525–541.

Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). Fog computing and its role in the internet of things. In *Proceedings of the first edition of the MCC workshop on Mobile cloud computing* (pp. 13–16). ACM.

Botta, A., de Donato, W., Persico, V., & Pescapé, A. (2016). Integration of cloud computing and internet of things: a survey. *Future Generation Computer Systems*, *56*, 684–700.

Bowerman, B., Braverman, J., Taylor, J., Todosow, H., & Von Wimmersperg, U. (2000). The vision of a smart city. In *2nd International Life Extension Technology Workshop, Paris* (Vol. 28).

Cardwell, D. (1994). *The Fontana history of technology*. London; New York: Fontana.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., … Scholl, H. J. (2012). Understanding smart cities: An integrative framework. In *System Science (HICSS), 2012 45th Hawaii International Conference on* (pp. 2289–2297). IEEE.

Cobham PLC. (2011). *Cobham Tactical Communications & Surveillance* (Brochure). Dartmouth, Canada.

Cobham PLC. (2014). *Safe Cities* (Brochure).

Dainow, B. (2015). Key Dialectics in Cloud Services. *Computers & Society, ETHICOMP Special Issue*, 52–59. https://doi.org/10.1145/2874239.2874247

Derry, T., & Williams, T. (1993). *A short history of technology from the earliest times to AD 1900*. Oxford: Oxford Paperbacks.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4*(4), 429–452.

Filipponi, L., Vitaletti, A., Landi, G., Memeo, V., Laura, G., & Pucci, P. (2010). Smart city: An event driven architecture for monitoring public spaces with heterogeneous sensors. In *Sensor Technologies and Applications (SENSORCOMM), 2010 Fourth International Conference on* (pp. 281–286). IEEE.

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, *29*(7), 1645–1660.

Guo, B., Wang, Z., Zhang, D., Yu, Z., & Zhou, X. (2013). Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things. *Journal of Network and Computer Applications*, *36*(6), 1531–1539.

Hancke, G. P., Silva, B. de C. e, & Hancke, G. P., Jr. (2013). The Role of Advanced Sensing in Smart Cities. *Sensors*, *13*(1), 393. https://doi.org/10.3390/s130100393

Hartswood, M., Grimpe, B., Jirotka, M., & Anderson, S. (2014). Towards the ethical governance of smart society. In D. Miorandi, V. Maltese, M. Rovatsos, A. Nijholt, & J. Stewart (Eds.), *Social Collective Intelligence* (pp. 3–30). Springer.

Ikonen, V., Kanerva, M., Kouri, P., Stahl, B., & Wakunuma, K. (2010). *D.1.2. Emerging Technologies Report* (No. D.1.2). ETICA Project.

ITU-T. (2008). *Ubiquitous Sensor Networks (USN)* (ITU-T Technology Watch Briefing Report Series No. 4). International Telecommunications Union.

Jaisingh, K., El-Khatib, K., & Akalu, R. (2016). Paving the Way for Intelligent Transport Systems (ITS): Privacy Implications of Vehicle Infotainment and Telematics Systems. In *Proceedings of the 6th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications* (pp. 25–31). New York, NY, USA: ACM. https://doi.org/10.1145/2989275.2989283

Jin, J., Gubbi, J., Marusic, S., & Palaniswami, M. (2014). An information framework for creating a smart city through internet of things. *IEEE Internet of Things Journal*, *1*(2), 112–121.

Kim, R., & Kaplan, S. (2005). Co-Evolution in Information Systems Engagement: exploration, ambiguity and the emergence of order'. In *3rd Int. Conf. on Action in Language, Organisations and Information Systems* (pp. 166–180). Limerick, Ireland.

Kim, Y., Kim, M., & Lee, Y. J. (2008). COSMOS: A Middleware Platform for Sensor Networks and a U-healthcare Service. In *Proceedings of the 2008 ACM Symposium on Applied Computing* (pp. 512–513). New York, NY, USA: ACM. https://doi.org/10.1145/1363686.1363812

Kitchin, R. (2016). *Getting smarter about smart cities: Improving data privacy and data security*. Dublin, Ireland: Data Protection Unit, Department of the Taoiseach.

Komninos, N. (2006). The architecture of intelligent cities: Integrating human, collective and artificial intelligence to enhance knowledge and innovation. In *Intelligent Environments, 2006. IE 06. 2nd IET International Conference on* (Vol. 1, pp. 13–20). IET.

Komninos, N., Schaffers, H., & Pallot, M. (2011). Developing a policy roadmap for smart cities and the future internet. In *eChallenges e-2011 Conference Proceedings, IIMC International Information Management Corporation*. IMC International Information Management Corporation.

Lamb, R., & Kling, R. (2003). Reconceptualizing users as social actors in information systems research. *MIS Quarterly*, 197–236.

Langheinrich, M. (2001). Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. In G. Abowd, B. Brumitt, & S. Shafer (Eds.), *Proceedings of the Third International Conference on Ubiquitous Computing* (pp. 273–291). Atlanta, USA: Springer-Verlag. Retrieved from http://www.vs.inf.ethz.ch/publ/papers/privacy-principles.pdf

Latour, B. (2005). *Reassembling the social: an introduction to actor-network-theory*. Oxford ; New York: Oxford University Press.

Law, J. (1992). Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systemic Practice and Action Research*, *5*(4), 379–393.

Lo, B., Thiemjarus, S., King, R., & Yang, G.-Z. (2005). Body sensor network–a wireless sensor platform for pervasive healthcare monitoring. In *Adjunct Proceedings of the 3rd International Conference on Pervasive Computing*. Hagenberg, Austria: Springer.

Luhmann, N. (1986). The autopoiesis of social systems. *Sociocybernetic Paradoxes*, 172–192.

Martínez-Ballesté, A., Pérez-Martínez, P. A., & Solanas, A. (2013). The pursuit of citizens' privacy: a privacy-aware smart city is possible. *IEEE Communications Magazine*, *51*(6), 136–141.

Maturana, H. R. (1981). The organization of the living: a theory of the living organization. *Cybernetics Forum*, *X*(2–3), 14–23.

Nam, T., & Pardo, T. A. (2011). Conceptualizing smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times* (pp. 282–291). ACM.

Petrolo, R., Loscrì, V., & Mitton, N. (2015). Towards a smart city based on cloud of things, a survey on the smart city vision and paradigms. *Transactions on Emerging Telecommunications Technologies*. https://doi.org/10.1002/ett.2931

Pettersson, J., Presser, M., Vercher, J. B., Muñoz, L., Galache, J. A., Gómez, L. A. H., & Hernández-Muñoz, J. M. (2011). Smart cities at the forefront of the future internet. In *The Future Internet Assembly* (pp. 447–462). Springer.

Picon, A. (2015). *Smart Cities: a spatialised intelligence*. Chichester, West Sussex, UK ; Malden, MA: John Wiley & Sons.

Psyllidis, A. (2015). Ontology-based data integration from heterogeneous urban systems: A knowledge representation framework for smart cities. In *CUPUM 2014: Proceedings of the 14th International Conference on Computers in Urban Planning and Urban Management, Cambrigde, USA, 7-10 July 2015*. MIT.

Restore NV. (2016). About Us. Retrieved from https://www.restore.eu/en/homepage

Restore NV. (2017, January 23). Restore deploys Smart Appliances in world-leading Smart City Project. Retrieved from https://www.restore.eu/en/news/press-release/restore-deploys-smart-appliances-in-world-leading-smart-city-project

Service Systems Group. (2015a). *HAT briefing paper 1 : Engineering a market for personal data : the Hub-of-all-Things (HAT)* (Working Paper). Coventry: Warwick Manufacturing Group. Retrieved from http://wrap.warwick.ac.uk/65605/

Service Systems Group. (2015b). *HAT Briefing Paper 2 : The Hub-of-all-Things (HAT) economic model of the multi-sided market platform and ecosystem* (Working Paper). Coventry: Warwick Manufacturing Group. Retrieved from http://wrap.warwick.ac.uk/65607/

Smart, P., Simperl, E., & Shadbolt, N. (2014). A taxonomic framework for social machines. In E. Mordini, V. Maltese, M. Rovatsos, A. Nijholt, & J. Stewart (Eds.), *Social Collective Intelligence* (pp. 51–85). Springer.

Söderström, O., Paasche, T., & Klauser, F. (2014). Smart cities as corporate storytelling. *City*, *18*(3), 307–320.

Soraker, J. H., & Brey, P. (2007). Ambient intelligence and problems with inferring desires from behaviour. *International Review of Information Ethics*, *8*, 7–12.

State Council of People's Republic of China. (2014). *Planning Outline for the Construction of a Social Credit System (2014-2020)* (Planning Outline No. GF No. (2014)21). China: State Council of People's Republic of China. Retrieved from https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/

Tabak, E. (2015). *Information cosmopolitics: an actor-network theory approach to information practices*. Waltham, MA: Chandos Publishing.

Taylor, N. (2005). *Urban Planning Theory Since 1945*. London: SAGE Publications.

Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems*, *5*(4), 187–196.

Varela, F. G., Maturana, H. R., & Uribe, R. (1981). Autopoiesis: the organization of living systems, its characterization and a model. *Cybernetics Forum*, *X*(2–3), 7–13.

Vinod Kumar, T. M. (Ed.). (2015). *E-Governance for Smart Cities*. Singapore: Springer Singapore. Retrieved from http://link.springer.com/10.1007/978-981-287-287-6

Zygiaris, S. (2013). Smart city reference model: Assisting planners to conceptualize the building of smart city innovation ecosystems. *Journal of the Knowledge Economy*, *4*(2), 217–231.