AN OBJECT-ORIENTED APPROACH TO THE PRIVACY PROBLEMS POSED BY DIGITAL INFORMATION AND COMMUNICATION TECHNOLOGIES

Louise Whittaker

8804130h

A Research Report submitted to the Faculty of Humanities, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Arts, Applied Ethics for Professionals

22nd February 2016, Johannesburg

ABSTRACT

The advent of digital ICT has raised a range of privacy problems that previously did not occur, owing to the scope and volume of data that can be collected, as well as the processing capacity of the applications These digital privacy problems are arguably *not* easily addressed within any particular traditional macroethical framework. We may therefore need to find an alternative approach.

One such approach is proposed by Luciano Floridi, who has devised "Information Ethics" - a macroethics for the identification, clarification and solution of digital ethical issues. While IE is useful in that it highlights questions of digital agency, it will be demonstrated that it is flawed when applied to problems of privacy posed by digital ICT. IE, however, points us in the right direction: An object-oriented ethics may be able to address the issue of digital agents.

In this essay I develop an argument for the moral intentionality of digital agents, based on the concepts of emergent value and indirect intentionality, that can underpin an object-oriented ethical approach to digital privacy for both digital and human agents. Using Nissenbaum's concept of contextual spheres, I provide normative guidelines for evaluating the competing interests of agent-objects in various digital spheres.

A brief evaluation of the approach, by way of an example, shows that the object-oriented LoA that I am proposing can be adopted for digital privacy problems. In such cases, and for the specific purpose of weighing up the competing rights and values of the agents and patients, we can treat all agents (human and non-human) as both intentional and real. This provides a reading of the case that goes beyond the consequentialist or ownership-based approaches, and arguably gets closer to the heart of the issue.

Where the approach is still open, however, is that we still have to justify and balance these interests. There is no simple formula to apply. A need for practical wisdom or Phronesis, in the form of a judicious weighing of moral interests, continues to apply to digital problems posed by ICT.

DECLARATION

I declare that this research report is my own unaided work. It is submitted for the degree of Master of Arts, Applied Ethics for Professionals, in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any other degree or examination in any other university.

_____

LOUISE WHITTAKER

22nd day of February, 2016

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my supervisor, Dr Robert Kowalenko, whose initial lectures on Information Ethics provided the impetus for this report. His comments on my work have never been less than absolutely rigorous, and he has been very patient with my delayed submission. His input has undoubtedly improved the final product greatly, although all remaining errors are, of course, my own.

Dr Brian Penrose runs an absolutely outstanding Programme in Applied Ethics, which has been the highlight of my varied studies. I enjoyed every minute of the programme, and was stimulated and extended as much as I had hoped to be.

My thanks also to the University of the Witwatersrand for the staff bursary that I received for a portion of my registration on the programme, which enabled me to pursue these studies.

DEDICATION

This research report is dedicated to my husband Carlyle, and my darling children Alexandra and William, who put up with my absences over weekends, and short temper while grappling with philosophical conundrums, with good grace.

TABLE OF CONTENTS

Abstract i

Declaration ii

Acknowledgement iii

Dedication iv

Introduction 1

Privacy Problems Posed by Digital ICT 4

Consequentialist and Deontological Approaches and Why They Can't Consistently Address Digital Privacy Problems 7

Information Ethics and its Limitations as a Macroethical approach to Digital Privacy Problems 12

Object-Orientation and Embedded Moral Value - or Why an Object-Oriented Approach is Nonetheless Useful. 21

Embedded and Emergent Values 22

Moral Agency 29

Intentionality 32

An Object-oriented Level of Abstraction. 37

Applying an object-oriented LoA to digital privacy problems 40

Conclusion 47

References 49

Introduction.

Privacy, secrets and spies remain very much in the news. The reverberations of the Wikileaks diplomatic communications leak continue around the globe. Legal actions continue to flow from Edward Snowden's exposure of project PRISM - the secret NSA monitoring programme that records global online and voice data. Far more commonplace - and yet therefore more concerning - is that profile data of a range of types of persons, from "new parents" to "people with cancer", is routinely bought and sold by corporations (Steel, 2013); Facebook continues to retain data that members have deleted and monitor members through the Like buttons embedded into third-party webpages (BBC News, 6 August 2014) and Uber has a "God View" function that allows some employees to track the location of customers who have used its car service (Rebeiro, 2016). What all of these developments have in common is digital information and communication technologies (ICT).

It seems that the advent of digital ICT has raised a range of privacy problems that previously did not occur, owing to the scope and volume of data that can be collected, as well as to the capacity of digital ICT applications to process attributive data so as to generate referential information profiles of individuals. Data that we consider suitable for public access can be processed to reveal things about us that we might consider private. This raises the question of whether there are morally significant privacy problems posed by digital ICT that should be resolved by recourse to a macroethical framework specifically designed for this task.

Privacy problems posed by digital ICT are arguably *not* easily addressed within any particular traditional macroethical framework. For example, it is hard to argue that there are necessarily or even possibly bad consequences for me if a government or corporation tracks my behaviour online. Nonetheless, the idea of being traced in this way makes many people, at some point, uncomfortable, particularly where the tracking is done in order to infer details about them that they may not have chosen to disclose publicly. This suggests that intuitively there is a moral problem with the practice. On the other hand, deontological or rights-based theories seem to rely on a concept of ownership of data, but it is impossible to argue that one owns, exclusively,

information about oneself that is already in the public domain.

The challenges posed by ICT seem to leave us with an ethical and policy vacuum. It can be difficult, using our existing theories, to reach a reflective equilibrium in which we can balance our acceptance of broad ethical imperatives with our moral intuitions in particular cases about the privacy problems posed by digital ICT. It might, of course, be that our moral intuitions will evolve, and that privacy simply isn't a problem at all. If however, that is the case, then we might have to accept, logically, that there is no aspect of ourselves and our lives that should not be disclosed to the public gaze. As in the dystopian novel, "Blind Faith", by Ben Elton (2007), we will all shortly post our sex, birth and death videos online, and go about more or less naked in public too. In such a world it is not just the personal that is intruded upon, but quite feasibly the political as well. The hero of "Blind Faith" is in the end, after all, executed for his religious and political non-conformity[1].

If the question of privacy remains important, then we need to find a consistent way to assess our moral intuitions about what is morally permissible and impermissible when it comes to breaches of our privacy by governments, corporations, and other individuals online. In other words, we need an alternative approach with which we can fill the vacuum.

One such approach is proposed by Luciano Floridi, who has devised "Information Ethics" - a macroethics that, he suggests, provides a conceptual ground, on the basis of which ethical issues posed by the advent of digital ICTs may be "more easily identified, clarified, and solved" (Floridi, 2008a, p.19). Information Ethics (IE) is a kind of environmental ethics where, rather than privileging the interests of humans (as in classical, anthropocentric theories), we consider the interests of the environment as a whole. Furthermore, in considering the rightness or wrongness of an action, we do not consider not only the moral well being of the agent, but the well being of the patient – the object of the action - which in IE is always a piece of information. While

[1] His non-conformity starts with his choice to vaccinate his child. If this seems like is an unlikely source of opprobrium, the recent fuss over Mark Zuckerberg choosing to vaccinate his daughter gives pause for thought.(https://www.washingtonpost.com/news/morning-mix/wp/2016/01/11/mark-zuckerberg-angers-anti-vaxxers-with-photo-of-baby-at-doctors-office-getting-vaccinations/)

IE is useful in that it highlights questions of informational versus psychological privacy and of digital agency, it will be demonstrated that it is flawed when applied to problems of privacy posed by digital ICT. IE, however, points us in the right direction: An object-oriented ethics may be able to address the issue of digital agents. By object-oriented I refer (following Floridi, 1999 p.49) to an applied ethics which is

- not anthropocentric or bio-centric but being- or onto-centric;
- neither only agent oriented nor only patient oriented, but oriented to both - which is what Floridi terms "object oriented";
- "not an ethics of virtue, happiness or duty, but of respect and care".

Digital ICT applications may demonstrate a degree of intentionality that is not necessarily equivalent to human intentionality but which nonetheless poses a question of moral import: Do digital ICTs have potential moral intentionality, and if they do, does that mean that they have rights and responsibilities – and particularly, for the purposes of this research, rights and responsibilities with respect to privacy? Do we need to consider the privacy rights of both digital agents and other (human) moral agents?

I would suggest that the answer to these questions is yes. In this essay I will develop an argument for the moral intentionality of digital agents that can underpin an object-oriented ethical approach to digital privacy for both digital and human agents, and explicate further what this object-oriented approach looks like. By drawing on Nissenbaum's (2004) concept of contextual spheres, I will then provide normative guidelines for evaluating the competing interests of agent-objects in various digital spheres.

What I am *not* attempting to do here is to provide a meta-ethics (which is what Floridi 1999) does do). I am rather developing an approach that can be applied to a particular domain (that of digital privacy), based on the construct of a Level of Abstraction (as defined by Floridi, 2002), and using, in addition, the concept of contextual spheres[2].

[2] Of course, without a prior knowledge of Floridi and Nissenbuam, these concepts are fairly cryptic, but In the essay that follows I hope to explain, critique and finally apply them.

In order to develop the argument, the essay will address each of the following questions in the sections that follow:

1. What sort of privacy problems are posed by digital ICT?
2. Why can't traditional macroethical approaches address these problems for us?
3. Why isn't IE an appropriate theoretical approach?
4. Why is an object-oriented approach nonetheless useful?
5. How can an object-oriented approach be applied to digital privacy problems?

Privacy Problems Posed by Digital ICT

Let us consider more carefully why digital ICT poses particular privacy problems. For readability, I will henceforth refer to the privacy problems posed by digital ICT as "digital privacy problems". What kinds of problems are these, and why are they of moral consequence?

Floridi (1999, p.52) describes four types of privacy, viz. physical, mental (or psychological), decisional and informational privacy. Physical privacy corresponds to freedom from sensory interference or intrusion. This is privacy of the body, essentially, and relates to what others can do to or with my body, whether through observation (where a peeping Tom would be an intrusion) or more directly (for example through a body search or dog sniffers at an airport (Bonfanti, 2014)). Mental privacy is freedom from psychological interference or intrusion – harm to my perceptions of myself and my relation to the world around me. For example, an intrusion that reveals aspects of myself that I would rather keep private - perhaps a coarse sense of humour, or a propensity to boast in private - may well affect my "identity, self-hood, thoughts, and so forth" (Tavani, 2008, p.162). Decisional privacy is freedom from procedural interference or intrusion – such as undue influence on how I make decisions regarding, for example education, health care, voting choices or so on. Information privacy is freedom from epistemic interference or intrusion – where facts about me that I would prefer to be private become public. The first three kinds of privacy relate to what others can do *to* me, whereas informational privacy relates to what others know *about* me.

Floridi (1999) argues that only informational privacy is relevant in the context of computer ethics[3], because it is only information that can be stored on digital ICT, and therefore only information that can be interfered with. However, a consideration of the value of privacy, and the relationship between these various kinds of privacy demonstrates quite quickly that they are all of concern in the context of digital ICT.

Privacy in itself is arguably not an intrinsic or core value (Moor, 1997, p.29). But it is certainly instrumental in ensuring some intrinsic values, because a violation of privacy can disturb, for example our freedom (if decisional privacy is breached), our autonomy (if decisional or mental privacy is breached), our capacity for human relationships (if mental privacy is breached), or even our physical safety (if physical privacy is breached). Informational privacy can be said to be important because it in turn prevents violations of decisional, mental or physical privacy, as it may be through intrusions on information about me that a party may impinge on my ability to make decisions, my autonomy and my bodily security or integrity. In other words, there is harm or disvalue in an epistemic intrusion to the extent to which it intrudes or interferes mentally, psychologically or physically. It is debatable whether informational privacy breaches themselves are violations, although Moor (1997, p.29) does argue that an informational privacy breach is in itself a breach of security, which, being necessary for the survival of a culture, is a core value.

It is in order to protect intrinsic values, through the instrumental value of privacy, that we have always protected personal information, whether through walls and curtains, private correspondence, secrets, confidential relationships and so on. The need to protect informational privacy is not particular to computer ethics or digital ICT. Can we then say that there are privacy problems that are particular to digital ICT, that deserve specific ethical attention at all?

There are at least two ways in which digital ICT can be said to exacerbate privacy

[3] Floridi refers to the field of ethical study in relation to digital ICT as Computer Ethics or CE. This term is somewhat dated as the ethical implications of computers are today wider than only computers or computer networks but include communication technologies, the internet, social media and so on as well – as indicated by the term ICT.

issues. These relate to the volume of data and the type of data that the use of technology creates. Firstly, the sheer volume of data about individuals that digital ICT now generates, transmits and stores has increased the potential for privacy violations to levels that were inconceivable even a few years ago. As an example, the activist Max Schrems has sued Facebook to force the company to reveal what data it was holding about him. He discovered 1222 pages of information, including information he had deleted or not consented to being shared (BBC News, 2014). Digital information has a quality of permanence (Nissenbaum, 1998, p.562), which is often not obvious, or even counter-intuitive to the subjects thereof. Because it is so easy to delete or remove data from the platform on which it is presented to us – and some programmes such as Snapchat even delete it automatically and almost instantly – it is easy to assume that it has in fact been deleted, whereas it may well be, and often is, more permanently stored on databases that are not visible to us. Although we might be vaguely aware of it, we don't often stop to consider that a retail store, for example might have, in addition to shopping data, demographic information such as the following:

> your age, whether you are married and have kids, which part of town you live in, how long it takes you to drive to the store, your estimated salary, whether you've moved recently, what credit cards you carry in your wallet and what Web sites you visit…your ethnicity, job history, the magazines you read, if you've ever declared bankruptcy or got divorced, the year you bought (or lost) your house, where you went to college, what kinds of topics you talk about online, whether you prefer certain brands of coffee, paper towels, cereal or applesauce, your political leanings, reading habits, charitable giving and the number of cars you own. (Duhigg, 2012 online)

Secondly, because of the amount of data, and the capacity of digital ICTs to process it, data that in another form or context would be attributive (about a person generically or persons in general), can be processed very rapidly to become referential (specifically referring to a particular person) – so that individually harmless information becomes identity-relevant, or capable of causing harm (van der Hoven 2008, pp.307-311). This has become particularly pertinent with the development of

"Big Data" predictive analytics systems, which can crunch data that is publicly available across a multiplicity of platforms, to be able to identify for example, when a particular person is pregnant, or ill, or about to retire. Facebook keeps 1222 pages of Max Schrem's data because it is expecting it to provide such potentially economically profitable insights. Nissenbaum (1998) refers to this as the problem of "privacy in public" – a "systematic relationship between privacy and information that is neither intimate nor sensitive and is drawn from public spheres" (p.559). The malleability and transportability of electronic records, together with their permanence, enable data surveillance and information harvesting that allow for informational privacy intrusions based entirely on public data.

In spite of these problems, it might be suggested that although the scale of the problem may be bigger, the essential type of problem is unchanged, and there are in fact no particular privacy problems posed by digital ICT. If that is the case, then our existing macroethical frameworks should suffice in dealing with the problems. Arguably however, this is not the case.

Consequentialist and Deontological Approaches and Why They Can't Consistently Address Digital Privacy Problems

Floridi (1999), in particular, has argued that is not possible simply to apply either deontological or consequentialist macroethical frameworks to digital ICT problems. According to him (2005, p. 193) there are two "popular" interpretations of privacy that rely on standard macroethical frameworks . On the "reductionist" interpretation, privacy is important because it prevents net undesirable consequences that might be caused by informational intrusion or interference. In other words there is a rule-consequentialist argument that privacy breaches in general have moral disvalue because they cause more harm than good. As in any consequentialist evaluation there is of course a potential for utility as well – for example a particular act of breaching a would-be terrorist's privacy is good for others' security – but the extent of the value or disvalue of a privacy breach is assumed to be estimable. On the "ownership-based" interpretation, privacy is important because information about a person belongs to a person, and he/she has a right to control it. This is a rights-based,

deontologial view, according to which the freedom, autonomy, security and other basic rights of the rational individual as an end in himself are to be protected.

If either of these views can be successfully consistently applied, then privacy problems posed by ICT, along with other ethical ICT (or Computer Ethics) issues, do not pose any unique sort of problem. According to Floridi however, "when consistently applied, both Consequentialism… and Deontologism show themselves unable to accommodate CE problems easily, and in the end may well be inadequate" (1999, p.39).

Floridi (1999) enumerates a number of reasons for which he considers it difficult to demonstrate consequential disvalue (or value) in ICT cases. These include: the fact that there may be no perceptible effects of, for example, Facebook storing 1000s of pages of my data without my knowledge; a perception that only "real life" has any moral consequence and that therefore "virtual" or ICT-based actions do not; and the impossibility of calculation of the actual consequences, given the "constantly changing infosphere" (p.40).

All of these particular criticisms amount to an argument that, in ICT cases, it may be difficult or even impossible to calculate consequences. This epistemological problem applies of course, to potentially all moral problems - how do we really know what the consequences of any action will be? For consequentialists however, such knowledge is not necessary for the purposes of making moral decisions, because the principle of maximisation of utility (good consequences), is a criterion for whether the action is morally right, rather than a decision process or procedure in the first place. As Sinnot-Armstrong illustrates - "just as the laws of physics govern golf ball flight, but golfers need not calculate physical forces while planning shots; so overall utility can determine which decisions are morally right, even if agents need not calculate utilities while making decisions" (2014, pp. 14-15). It is probably adequate rather, for decision-making purposes, to follow our moral intuitions, which intuitively take account of the most important expected consequences. We can, after all "have strong reasons to believe that certain acts reduce utility, even if we have not yet inspected or predicted every consequence of those acts" (Sinnott-Armstrong, 2014, p. 16). Where we really cannot assess the consequences, it simply points to the "severe

limits to our knowledge of what is morally right", which is not to be unexpected in moral problems in general, and therefore logically in ICT-based problems too.

A further problem for consequentialism that Floridi (1999) identifies in ICT cases is the potential for a diminished sense of responsibility or accountability on the part of agents (p.40). In the case of privacy issues that would amount to a diminished sense of responsibility on the part of the persons responsible for deciding to keep my data – given that they don't know me at all. This certainly might explain why agents are more inclined to act unethically in certain circumstances rather than others, but it does not, in consequentialist terms, justify their actions. As Singer (1972) famously points out, we are as morally obligated toward the distant as we are towards those who are close to us, in proximity or relationship (p. 232).

With regards to the "rights-based view", Floridi (1999) argues that Deontology is too "inflexible" for ICT problems, and that there are too many conflicting "rights, duties and moral values" for Deontological approaches to be consistently applied (p.39). Moreover, it can be difficult to perceive of these rights as carrying any serious weight, given the "ludic nature" of many digital interactions (Floridi, 1999 p.40). Again, these are problems that can be raised against Deontological approaches in general (a discussion of which is beyond the scope of this essay). Suffice it to say that perhaps an expectation that any macroethical framework could "accommodate CE problems easily" (p. 39) is unrealistic.

Nonetheless, of Floridi's (1999) objections, the following is probably the strongest: He suggests that the anthropocentric nature of Deontology makes it unsuitable for analysing situations in which non-humans (that is, computers) are implicated as the possible initiator of action (p.40). In fact this objection applies to both consequentialist and deontological approaches. What is the moral disvalue of Google automatically scanning my emails so that its servers can send information about me to other servers? How would anyone calculate the net value of that action? What rights or responsibilities do emails and servers have anyway? And even if I have a right to my own emails or Facebook posts, once information has been generated from base data about me, I don't own that information, so can I have any rights in respect of it?

The problem with the classic or standard macro ethical approaches is not so much that they are strained by CE cases when they are applied, but that they cannot begin to be applied to many cases - they cannot account for how we might begin to hold non-human agents accountable, or concern ourselves with the consequences for - or rights of - non-human patients.

This then raises the following question – if consequentialism and deontology can't account for the moral import of digital privacy, then does it have any moral value at all? Or should we simply concede that there are really seldom significant consequences or any real breach of rights in breaches of privacy in the digital realm? Certainly policy differences between the United States and Europe, and the casual attitude of most teenagers towards social media seem to point to a spectrum of views on this – the logical end point of which will be that the age of privacy is, as Mark Zuckerberg would have it, over (Kirkpatrick, 2010).

The problem that we would then face, however, is what a completely transparent world would actually look like. In a completely transparent world, no information about a person, no communication, no detail of their likes, dislikes, personal or political preferences, possessions or behaviour would not be made visible to anyone who cared to see. As pointed out by Rachels (1975), this would be a world in which it would become impossible to form varied relationships with different people – I would be as transparent to the passenger sitting next to me on the bus as I am to my closest friend. In such a situation there would be no such thing as an intimate bond with any one, and our capacity for human relationships would be severely diminished. Subject to the relentless gaze of every and any other, our autonomy is likely to be severely impeded (Johnson, 1994 cited in Moor, 1997), our ability to "be one's own person" (Christman, 2015, p.1) by definition overtaken by the fact that we are now everybody else's.

It seems unlikely that anybody would want to live in a completely transparent world. Therefore we ought to have some principles for privacy in the real world. This is precisely what we do in our everyday practice, where, except in the most oppressive societies, we have the right to speak or not, to have confidential relationships of both

personal and professional kinds, to cover ourselves up or reveal ourselves to a great or lesser extent as we please. This is not to say that anyone has complete privacy any more than we have complete autonomy or freedom in the real world, but certainly there are degrees of privacy that we take for granted and seem to value.

Undoubtedly, in the twenty-first century, digital ICT has potential effects on our existence in the real world. These effects are obvious when we interact with individuals online. Cyberbullying, for example, even if it is entirely virtual, has effects on the real psyches of real people and has lead to actual suicides (Megan Meier Foundation, 2015). There are also real world effects of the use of information by corporations. For example Target (an American chain store) uses data mining to identify customers who are pregnant, and in one case sent coupons for baby clothes to a teenager, whose family did not know she was pregnant at all. Even adult women don't necessarily like the idea that a corporation is tracking their reproductive status (Duhigg, 2012) and so Target does not disclose to customers that it is doing so, or even make it obvious that it has – it sends coupons for lawn mowers along with coupons for nappies, so that the target marketing is not obvious.

Therefore as much as we need privacy in the real world, we also need principles for digital privacy. If consequentialism and deontology in their classical forms can't provide principles that work for digital privacy then we have to find other principles that will apply in this context. In an attempt to do just that, Floridi has developed "Information Ethics" (IE) - a macroethics that, he suggests, provides a conceptual ground, on the basis of which ethical issues posed by the advent of digital information and communication technologies (ICTs) may be "more easily identified, clarified, and solved" (Floridi, 2008a, p.19). One of the issues that he has analyzed is informational privacy, and in particular the informational privacy problems posed by digital ICT (Floridi, 1999 and 2005).

Information Ethics and its Limitations as a Macroethical approach to Digital Privacy Problems

In this section I will suggest that IE does not provide us with an adequate means of drawing moral distinctions between entities, and that therefore the moral laws of IE are difficult to apply in practice. With specific reference to informational privacy, the distinction between the information object, and information *about* the object, seems to create similar difficulties for using IE as an appropriate theoretical approach to digital privacy problems.

Floridi (1999), having identified the problems with reductionist (consequentialist) and ownership-based (deontological) approaches, proposes a theory that is neither anthropocentric, nor agent-oriented, but rather "infocentric, and object-oriented" (p. 43). As mentioned in the introduction, IE is a kind of environmental ethics where, rather than privileging the interests of humans (as in classical, anthropocentric theories), we consider the interests of the environment as a whole. Furthermore, in considering the rightness or wrongness of an action, we do not only consider the moral well being of the agent, but also the well being of the patient – the object of the action. Whereas environmental ethics is bio-centric and patient oriented, Floridi suggests that in order to deal with ICTs we need to abstract further to be infocentric and object oriented.

This is done by adopting a theoretical level of abstraction (LoA) at which all entities are considered to be "consistent packet[s] of information" (p. 43) comprising the properties of the object and the possible states and behaviours of the object in response to stimuli[4]. All processes, at this level, "can be treated as information processes" (p.43). The totally of entities and processes (i.e. the environment, or everything) is the infosphere. An agent is "any entity capable of producing information phenomena that can affect the infosphere" (p.44). Entropy is any "decrease or decay" of information leading to an absence of content or pattern in the

[4] The relevant object, thus defined, would possibly then need to consist of infinite packets of information, since possible responses are not necessarily finite (Thanks to Dr Kowalenko for pointing this out). This is consistent with the idea that the totality of entities (including their processes) equates to the infosphere, since the infosphere is in fact everything.

infosphere (p.44), that is, “any process that negatively affects the whole infosphere” (p.45)[5].

This descriptive metaphysical account of the infosphere then becomes basis for a macroethics, as Floridi suggests that “even more fundamental[y] than life and pain” (1999, p. 45), *being*, which is now understood as information, “has an intrinsic worthiness” (p.45). As such welfare is determined in terms of existence. Any information entity (any thing) has a “right to persist in its own status” as well as a “right to flourish” p.45). Being/information *is* good and non-being/information entropy *is* evil.

There are thus four moral laws in IE as follows:

1. entropy ought not to be caused in the infosphere
2. entropy ought to be prevented in the infosphere
3. entropy ought to be removed from the infosphere
4. information welfare ought to be promoted by extending, improving and enriching the infosphere (p. 47).

Information Ethics is, however, applied by Floridi (1999, 2005) to the question of informational privacy not through an application of these laws but through a consideration of two ontological assertions that follow from the infocentric LoA.

The first of these is the assertion that an object and its information are, in the infosphere, one and the same thing. Therefore a person *is* her information. An instance of epistemic interference or intrusion (a violation of informational privacy) that occurs because another agent obtains information about the person is therefore a violation of her “integrity and uniqueness as an informational entity” (Floridi, 2005, p. 195) – her very person. Informational privacy violations therefore are akin to kidnapping, rather than theft - which is how they are traditionally understood, if privacy rights derive from the “ownership” of one’s personal information.

[5] Floridi’s definition of entropy is something of a combination of the information theory definition, where entropy is “a measure of the loss of information in a transmitted signal or message”, and the definition as applied in thermodynamics, viz. a “measure of randomness” or “disorder” (http://dictionary.reference.com/browse/entropy).

Informational privacy, according to IE, is therefore a "fundamental and inalienable" (p. 195) right[6].

The second ontological assertion is that privacy is a function of ontological friction. Ontological friction is defined as "the forces that oppose the information flow within (a region of) the infosphere" (Floridi, 2005 p.186). The lower the level of ontological friction, the smaller the information gap between entities, and the lower the level informational privacy possible. In the digital world, level of ontological friction is mostly significantly reduced, because the infosphere is being "re-ontologized" (p.188) as both information and processes are digitized.

Re-ontologization is the process by which "everything" is moving into the digital realm. Floridi (2005, pp.188:189) points out five ways in which this is happening: Firstly, the sheer volume of digital information is vastly increased and increasing. Information that was previously stored in analogue form is being converted to digital form through both private and public projects, Google Books being an example of the latter. Furthermore, much information is today only produced in digital form; Secondly, there is an increasing homogenization of processor and processed information when both are digital - that is, computers and microchips are producing information and also processing and analyzing it, a development which is rapidly causing the Internet to become an Internet of Things, which could contain 50 billion objects by 2020 (Evans, 2011). Thus, as an every increasing range of types of objects (fridges, cars, watches) have both increased computational power and internet connections, they are capable of recording, processing and sharing information about themselves, so that their digital presence is as relevant as their physical one; Thirdly, there is an evolution of new informational agents as humans become ever more attached (physically and metaphorically) to their enhancing personal ICTs. The teenager lost without her iPhone is a clichéd example, the diabetic with an automatic insulin pump that closely mimics a healthy pancreatic function is a more serious one; Fourthly, where information and processors are all digital, all interactions are "informationalized" into read/write or execute instructions. In other words, only

[6] Floridi applies IE only to informational privacy.

digital actions can be taken by digital agents on digital information; And fifthly there is a mutation of old agents into informational agents as augmenting technologies draw us into the digital world, where human beings can be "present" in a digital form, having entered the digital world through a multiplicity of interfaces, gateways, portals and so on.

All of this, Floridi suggests, means that there is currently an "unprecedented migration of Humanity from its Umwelt to the infosphere itself" (2005, p.189). In this re-ontologized digital infosphere - where the level of ontological friction is so greatly reduced by digital volume, homogenization, evaluation, informationalization and mutation - privacy problems are greatly exacerbated. Information about an individual becomes so freely available that the best way for an individual to protect her personal self, in the form of her information, is to keep her identity data as close as possible to the physical manifestation of her being – through biometric identification (p.198). An identity thief (which is what a privacy intruder is) would presumably then have to kidnap her physical person to steal her identity, rather than just her data.

The question posed here is whether IE constitutes an appropriate theoretical approach to digital privacy problems. As Floridi (2008a) points out, IE - being a foundationalist project - "is not immediately useful to solve specific ethical problems (including computer ethics problems)" (p.19). Nonetheless, it seems reasonable to assume that IE would have something to say about privacy (Stahl, 2008, p.101), since privacy is "directly linked to information and access to information" (Stahl, 2008, p. 101), and of course Floridi has himself applied it thus (Floridi, 1999, 2005, 2008b). In considering the question we need first of all to determine whether information ethics provides an appropriate theoretical approach to moral judgements in general, and then to examine how it informs considerations of privacy in particular.

The basic premise of IE is that Being, understood as information, has an intrinsic worthiness. Non-being, entropy, is evil. As such, any entropy should be prevented. Because the entity and the information that describes that entity's properties and possible behaviours/actions/states are one and the same thing, entropy necessarily refers to the breakdown or destruction of the thing itself, whether it is abstract or

physical. This is clear from the example Floridi (1999) gives of the boy who is vandalizing cars – his "game is only increasing the level of entropy in the dumping ground….he ought to stop destroying bits of the infosphere" (p.54).

Because IE a. abstracts to the level of the infosphere, and b. says that entropy ought to be prevented, not to be caused and removed from the infosphere, it therefore says that any form of destruction ought not to occur. That seems to imply that all entities are equivalent (because they are all information), and that none must be destroyed. This had led Brey (2008), for example, to conclude that "IE tells us that we should be equally protective of human beings and vats of toxic waste, or of any other information object" (p.112).

Floridi's (2008b) response to this objection is to note that IE does not suggest "some daft idea about the intrinsic value of Shakespeare versus Dan Brown" but rather asks us to consider all information as "minimally and overridably" ethically valuable in itself, "*to begin with*" (p.193*)*. We must then have some "general, basic and robust principles in place" (p.194) to help us to decide that human beings are more valuable than vats of toxic waste, or Shakespeare more valuable than Dan Brown. The problem is, that at the infocentric LoA we don't have such principles. We have only the principles pertaining to entropy, and they don't distinguish between entities at all.

Perhaps the entropy principles themselves can help us value competing entities. If toxic waste destroys other entities then I suppose that the infosphere as a whole would flourish if the toxic waste were destroyed. There are two potential problems then. The first one is how to calculate the overall level of entropy in the infosphere, given that course of action. This problem does not seem any different from the consequentialist problem of having to calculate the utility of an action – it's just at a different LoA (Stahl, 2008, p.104). In that case it is not clear exactly what abstracting to this level has contributed. The second problem is what we would do if two actions - one ethical and one not - were roughly equal in terms of their respective effect on entropy. Without already knowing which action was ethical we would have no way to proceed. We are back to assessing "Shakespeare versus Dan Brown" in a way that might lead us to destroy something worthwhile to us.

Alternatively, if no such calculation is necessary because the minimally and overridably valuable entity is to be overridden "in view of moral concerns formulated by other macroethical analysis at lower LoA" (Floridi, 2002 p.302), then it is not entirely clear what is gained by abstracting all the way to the level of the infosphere, only to bounce back down again, as it were.

Floridi suggests that what is gained is that "IE has its own special field of application, CE" (2002, p.302), and that that "IE has already been fruitfully applied to deal with … the problem of privacy" (2008a, p.19). However, the fact that Floridi himself does not seem to apply the four moral laws directly to the question of privacy[7], seems to suggest that the notion of entropy has limited usefulness when considering informational privacy. It may even be directly contradictory – if entropy is to be prevented and information welfare to be extended (laws 0 and 3), but information flow (which is actually a process of copying information, when it is digital) needs to be restricted for privacy, then privacy measures contradict the moral laws. Free information flow and a complete absence of privacy would be a good thing. In this case, for example, Max Schrems would have no justifiable concern about Facebook keeping his data. In fact Facebook would be morally obliged to do so, since to delete the data would be to increase the loss of data, or entropy, which is evil.

This leads me to consider whether IE helps us to analyse questions of ICTs and privacy. That is, is it particularly appropriate when applied to entities which are themselves digital items of information (understood in the regular sense of the word), and therefore to the protection of digital information?

An information object is any entity, at the level of its information. That is, the object and its information are conceptually one and the same thing. There is an important distinction here, however, between the "information object" and what we normally understand as information about that information object. Let us call the latter information-*about*. Consider that information-*about* is generally understood to be

[7] I refer here to the papers cited.

data in some recorded form that is meaningful in respect of some entity. Until the data is recorded – say in a photograph, or words in a letter, or details on a form, or a record in a database, information-*about* simply doesn't exist[8]. Once it does exist, it is then an information object itself – that is, it has its own properties and possible states[9]. So now that object (the photograph, the form, the data entry) is now *its* information. But can we logically say that the second object *is* the first object? Clearly we can't. They are two separate, although closely related, information objects.

The problem is that Floridi (1999, 2005) implies that we should say that they are the same when he discusses digital information, and privacy in particular. Floridi is concerned with informational privacy - "freedom from epistemic interference or intrusion, achieved thanks to a restriction on *facts about* S that are unknown or knowable" (Floridi, 1999, p.52, my italics). He says that if someone's privacy is breached because information-*about* them is interfered with, then that is akin to kidnapping because the person *is* her information. "Cloned information… is a part of the observed herself" (Floridi, 2005, p.195). Therefore, the nature of the information breach is irrelevant because it is a breach of the person herself - the "packet of information" (Floridi, 1999, p.53) that she is. For Floridi privacy is nothing less than the defence of the self, because "*any* information about ourselves is an integral part of ourselves" (p.53) - and at the infocentric LoA he means this literally and not metaphorically.

I would argue, however, that the person is only *her* own semantic information, even at the level of the information object. She is not, even at the infocentric LoA, identical to the information-*about* herself. Common sense tells us that information *about* ourselves may not even be correct – it may be outdated or simply erroneous - in which case, for Floridi, it is not actually information at all. Even if we were able to

---

[8] The exception to this may seem be direct observation. However, even a human being directly observing another human being forms perceptual beliefs with a semantic content - information-*about* that person. If casual observation (in public spaces) necessarily allows us to access information that amounts to a violation of a person's integrity, then we all need to close our eyes (and ears).

[9] Even a record in a database corresponds to a particular physical configuration of atoms. It's not necessarily immediately obvious, but it is a physical object as such. And of course it doesn't need to be a physical object to be an information object.

upload the entirety of our information to a digital reader of some kind, it would be out of date within microseconds as our biological and psychological makeup are not fixed. Information-*about* ourselves simply cannot be co-existentially equivalent to ourselves.

It is intuitively appealing to understand that privacy is somehow an intrusion on our identity, and the closer the information-*about* is in relation to the information object, the more this is likely to be the case. A diary, in which I reveal my closest secrets - the ones that might otherwise only exist in the information object that is me - may reveal me to the reader as much as if she had been able to read my mind. So in that sense the idea of the information object is useful, because it might help us to identify the *kinds* of information that might warrant privacy protection. But an insistence that a person is her information, *and* that that particular existential information is itself intruded on by accessing information-*about* the person, simply can't be true.

The kinds of information-*about* a person that are closest to her identity (that is, the information object) are, it seems to me, the kinds of information that, if intruded upon might violate her mental or psychological privacy, because there is a clear sense in which some information (how I feel about my partner) can be more psychologically private than other information (the public fact that he/she is my partner). Tavani (2008) has noted that "based on what Floridi says…it would seem that his distinction between informational privacy and mental privacy breaks down" (p.162). I would suggest that an alternative interpretation would be to say that IE in fact does not account adequately for informational privacy, because it is actually directed to questions of identity, or psychological privacy. This is supported by Floridi's (2008b) comment: "in the sense that 'to be is to be an information entity'… there is a continuum between the informational nature [the being] of an individual and his or her mental/psychological privacy, …where one decides to draw a threshold…is probably a matter of circumstantial agreement" (p.199).

The challenge to IE here, is that it precisely does not deal with circumstantial or contextual matters, whether these be the question of whether mental privacy is intruded upon by a particular information-*about* intrusion, or whether a person ought reasonably, in the circumstance, to have to expected privacy in the first instance

(Tavani, 2008, p.163). The question, again, is how far we can explore issues of privacy at the infocentric LoA, if, in the end, we need to consider contextual realities anyway.

Furthermore, if IE has limitations in exploring issues of informational privacy in general, then it is not clear how it will be appropriate in exploring privacy problems posed by ICT in particular. These problems arise from the reduction in ontological friction that digital ICT produces, but this ontological friction relates to information-*about* information objects, and not necessarily the original information objects themselves.

A final objection to IE as a basis for a theory of privacy in the digital age is that Floridi seems to use the term "infosphere" in ambiguous ways depending on whether he is discussing the metaphysics of IE, or the question of informational privacy in the digital age.

On the one hand, IE specifies that the infosphere "is Being considered informationally, as simple as that" (Floridi, 2008, p.200). On the other hand, Floridi (2005) says that "the re-ontologization of the infosphere has been causing an epochal, unprecedented migration of humanity from its *Umwelt* to the infosphere itself" (p.189). Since the re-ontologization of the infosphere refers to processes of digitization, it does not seem unreasonable to suggest that the infosphere to which humanity is migrating is now the digital infosphere. The problem is that this somehow leaves the rest of the infosphere (the *Umwelt*) behind.

Take for example, the assertion that "in the re-ontologized infosphere …where there is not ontological difference between processors and processed, interactions …are all interpretable as read/write activities with 'execute' the remaining type of process" (Floridi, 2005, p.188). Such a re-ontologized infosphere simply doesn't include non-digital agents who do things such as hug their children. "Hug" is neither read nor write, and even if one were to suggest 'execute hug' as an interaction, that still requires 'hug' as a process itself, and one that requires a non-digital presence to boot. Perhaps it is not all that unreasonable of Capurro (2008) to suggest that in this

case the 'infosphere' is conceived as "separated from what phenomenologists call the 'life world'" (p.170).

What implications does this ambiguity have for Floridi's theory of informational privacy? I think that Floridi (2005) is overstating when he suggests that the infosphere is being re-ontologized. It simply isn't the case that processor and processed are now all digital, and until the entire physical world is both uploaded and physically erased it cannot be. What this means is that significant ontological friction exists between the information objects that take physical form (for example humans) and digital information objects. Anyone who has tried to access data in a format that is no longer supported by her computer can tell you this. Even if data are accessible they are often no longer current, and even if they are current, they are necessarily information objects in their own right – information *about* another entity, not actually that entity themselves. I suspect that a (probably overly confident) sense of this ontological friction might be what makes digital natives sanguine about the protection of their privacy online.

In conclusion, Floridi seeks to develop a macroethics, a theory that can underpin considerations of what we ought to do in particular circumstances. The theory relies on a methodological LoA that is itself coherent, but that is arguably not entirely helpful in addressing particular moral problems. With regard to questions of privacy, the distinction between the information objects and information-*about* objects detracts from the appropriateness of IE as a theoretical basis for considering digital privacy problems.

Object-Orientation and Embedded Moral Value - or Why an Object-Oriented Approach is Nonetheless Useful.

IE is, however, useful because it may point us in the right direction. That is, an object-oriented ethics may be useful if it addresses the issue of digital agents.

An object-oriented ethics would be one that is not human-agent centric, but one which adopts a level of abstraction at which all potential agents – human and digital –

are considered, without reducing them all to the level of information, but by considering them all as potential agents and patients. In this section I will develop an argument for the intentionality (not personhood[10]) of digital agents that can underpin an object-oriented ethical approach to digital privacy for both digital and human agents.

Embedded and Emergent Values

The first step in this argument is to establish whether digital agents are capable of producing consequences of value or disvalue which are not wholly predetermined when the programmes are developed in the first instance. In other words, do digital agents embody emergent values, independently of the intentions of the designer, if at all?

An influential paper regarding values and computer technology is "Values in technology and disclosive computer ethics", in which Philip Brey (2010) argues that computer technologies have embedded values. There is of course an extended literature on this topic, some of which I will reference in the discussion that follows, but it is useful to follow Brey's (2010) particular argument for embedded values.

Brey argues that technologies have built-in consequences, which he defines as "recurring consequences that manifest themselves" in "all central uses of the artifact" (2010, p. 44). These consequences might be intended by the designer (for example, that the saw cuts wood), or not (for example, that the car emits fumes), but if the technology tends to produce them, then, he argues, they are built into, or embedded in, the technology.

Brey then considers the question of value. He argues that "to find something valuable is to find it good in some way" and therefore that "values…correspond to idealized qualities or conditions in the world that people find good" (2010, p.46). In this sense

[10] To argue for personhood would require that digital agents are rational beings capable of self-perception and self-directedness. While one might argue that digital agents have the potential to be persons in this way, or that they are virtually so, this is not yet the case, and doesn't need to be for an argument in favour of moral intentionality to hold, as I hope to demonstrate.

“values” are abstract ideals such as justice, democracy, autonomy and so on. To the extent that the embedded consequences of a technology produce actual conditions or qualities that are good (or bad), therefore, the technology embeds the corresponding value.

This is, as Brey notes, a causalist conception of embedded values – the technology, in a recurring manner, through the central use of the technology, causes the good or bad condition to occur. As such, it makes real or present the corresponding value (2010, p. 47). Whether the value is embedded because of the pre-existing preferences of the designers, or because of technical constraints, or because of the way in which the system turns out to be used (Friedman and Nissenbaum, 1996), it is that which is caused - the outcome - that is a realization of the embedded values.

The question then arises as to whether computer technologies in particular are prone to recurring consequences that produce conditions of value or disvalue.
To the extent that computer technologies can operate autonomously – that is, without immediate operation by the user, Brey argues they are capable of engendering “their own consequences” (2010, p.45). That these consequences can be more or less good or bad - of value or disvalue - he illustrates by way of example. Systems that show bias, or undermine user autonomy, or limit freedom of information, or favour the powerful (2010, pp.48:49), have embedded disvalue in themselves, and therefore we need to evaluate the ethics of the technology itself, “largely or wholly independently of actual uses of the system” (2010, p.41).

Such embedded values are often, according to Brey not immediately evident. That is, the technology is morally opaque, rather than morally transparent. This is particularly the case for information technologies, where the operation of the technology is often at a distance , as well as technically complex, and more significantly, closed off from the user, or “black-boxed” (2010, p.51). So, while it is generally understood how hammers and guns and even technologies like ventilators function, it is often not understood by the user how computer hardware and software function. And if, for example, we don’t know the technical details of how facial recognition software operates, then we won’t know that it has a built-in technical tendency to recognize

people with darker skins, which means that it may have an embedded (although unintended) value of racial discrimination (Introna, 2005).

This argument - for built-in consequences of technology - is in opposition to the neutrality thesis, which states that there are "no inherent consequences to technological artefacts" (Brey, 2010 p.43). On such a thesis, a tool may be used for any purpose to which the user puts it. A hammer may be a murder weapon, a gun or a collector's item. Because the tool has no agency, it cannot be said "do" anything, and therefore the consequences of its use are not inherent in the tool, but in the use to which it is put. The responsibility for such use rests with the user. Therefore, there is no need to consider the tool itself at all; we need only concern ourselves with the practices of the user.

A focus on practice and use is generally associated with what Brey (1997) refers to as "strong social constructivism" - a position that says that technology is to be explained only through "reference to social practices" (Brey, 1997, p.61). On a strongly constructivist view, there is no such thing as a "real" object, independent of the social construction of that object. A piece of metal attached to a piece of wood is only a hammer because we have socially constructed it, through use, as such. Therefore there can be no real, inherent or intrinsic properties or consequences of the object. The object itself, as well as the consequences thereof, will always be the outcome of a process of interpretation and settlement – a "genuine social construction" (Brey, 1997, p.6) of reality. It is this position that leads "strong" social constructivists such as Grint and Woolgar (1995, p.306) to argue that

> [t]he politics and values of technology result from the gaze of the human; they do not lie in the gauze of the machine. . . . What the thing is, even what its exact capabilities and effects are, is not something that any kind of detached, objective, or realist analysis seems capable of constructing.

Such a strong view directs our attention entirely away from the properties of the technology (the gauze of the machine) to the use thereof. Thus, as Messerly (2007, p. 19) argues, "medical ethicists don't focus on the design of ventilators, but on the

practice of euthanasia". Most of what Brey (2000) refers to as "mainstream computer ethics" (p.10) does the same, focusing on issues of policy and use, and not on the design of the technology.

Such a strong view of the neutrality of technology, however, does not necessarily accord with our moral intuitions. We talk, for example, of "beating swords into ploughshares", because we understand that there is something about a sword as such that not only makes it unsuitable for ploughing, but also speaks to a potential for violence.

Is there a way to allow for the fact that technologies can be interpretively flexible (swords can be used decoratively or ceremoniously), but at the same time, have intrinsic properties? As Brey sets it up, we cannot give credence to the neutrality of technology as soon as we allow that some technologies have built-in consequences from their central uses. In such a case, we must focus on the sword as a weapon – on the technology itself. In Brey's own argument however, there is still a connection between use and consequence – the built-in consequence comes from the central use of the technology. He argues that as soon as a technology comes to be used in a particular way, it acquires particular built-in consequences. In fact "built-in" is completely misleading, as it implies that embedded consequences have to be there from the start, whereas Brey himself points to the fact that they can be emergent - that is: "not necessarily a reflection of the values of the designers", but arising when the "context in which the system is used is not the one intended" (p.50). Therefore, I don't see how Brey can get away from use to focus only on the technology – to propose an "ethics of computer systems separate from the ethics of *using* computer systems" (p.41 emphasis in original). But I also don't see why he needs to.

Only "strong social constructivists" completely disavow the properties of the technology itself as being relevant[11]. "Mild social constructivists" will consider how properties of a technology contribute to "social shaping" (Brey, 1997, p.6). And constructivists, such as Latour (1991), argue that a distinction between "social" and

[11] And "strong" social constructivism is a position that not many sociologists of technology will hold to these days in any case (thanks to Dr Kowalenko for this point).

"technical" is not helpful, and that human and non-human agents co-constitute technologies in the first place. The sword is a sword because it is sharp and strong and used for assailing people. Such an account does not discount use, any more than it discounts the properties designed into the tool.

Constructivists acknowledge that a tool is precisely a tool because it is designed with a particular task in mind. In order to be used successfully as such, it must engage the user in a particular way. Latour (1992, p.255) refers to this as the technology "enforcing its script" on the user, to produce a network (Latour, 1991, p.129)– human and non-human agents working together to effect an outcome. Of course, this isn't always the outcome that the designer had in mind – scripts can be rewritten or translated (Akrich and Latour, 1992 p.264). What we need to do, in order to properly evaluate the technology, is to consider the actor-network – the human/machine artefact as a whole. If we do this, then we can allow for flexibility in use, without discounting the embedded consequences of the technology. When we do this, we must understand that "technology" does not refer solely to the technical artifact, but to the human/machine artefact (what Haraway, 1991 calls a "cyborg"). Perhaps, although it really isn't clear, this is indeed what Brey (2010) means by "computer systems" (p.41). Indeed, arguably, he must mean this, as very few sociologists of technology would account for a computer system, except as a socio-technical system.

For the purposes of an object-oriented ethics, we can then conceive of this constructivist view as a Level of Abstraction at which all agents – human and non-human – operate. It is not even necessary to argue for the ontological "reality" of the actor-network – although Latour, being a realist (Latour, 1999, p.1), would strongly do so.

Assuming we can at this LoA allow for embedded - but almost always in some way emergent - consequences of technology, then does it follow that values are embedded in the technology? Brey does not give a particularly thorough account of what he means by "values". What we can discern from his account is that values reflect a certain conception of the good, that they are plural, and that they "correspond to idealized qualities or conditions in the world". While the correspondence of value to

ideas of "good" and "bad" (Schroeder, 2012), is probably uncontentious, value pluralism, and the choice of particular values without some sort of theoretical argument for them, are not.

However, holding to one side for the moment what we might consider as valuable or not, can we accept that some value or values might indeed be embedded and emergent in technologies? By this I mean that, de facto, values result from design, use and context of the technology, which are more or less stable at a point in time, although subject to interpretive flexibility (Orlikowski, 1992) and therefore translation and reinscription[12]. If we accept the notion that technologies have embedded and emergent consequences, as a results of scripts that are written into the design *and* rewritten or translated in the use of the technology, then it seems valid to suggest that we can evaluate these consequences to see if they are good/bad, better/worse (Schroeder, 2012). If the embedded consequence is good, then we can argue that it has value. If it is bad, we can argue it has disvalue. This value/disvalue can surely be said to be embedded in the technology.

I would suggest that we could not feasibly argue that technologies never have consequences that are arguably good or bad - we would then have to accept that technologies only have consequences that are always and un-controversially neutral. Messerly (2007) seems to think this is possible. He asks, incredulously it seems to me, "do people really believe their software was designed to bring about social change?" (p.21). But he is missing the point. Firstly, software is designed to bring about social change, because it is designed to change practice. One might find the change unobjectionable to the point that one doesn't even really notice it, but that doesn't mean it's not there. Secondly, there are enough examples of negative and positive outcomes to make it evident that the consequences of technologies are not neutral.

---

[12] This is not to suggest that there is unlimited interpretive flexibility, or that the technology is not ultimately grounded in a physical reality, which constrains which values can be embedded and emergent. But to focus *only* on that physical reality, or on the initial intention of the design, is to ignore the factual emergence of alternative scripts, which does actually occur. For examples of alternative scripts in the use of ATMs for example, see Introna, L.D. and Whittaker, L. (2006) Power, Cash, and Convenience: Translations in the Political Site of the ATM, *The Information Society*, 22, pp.325–340.

Therefore, technologies must have embedded values of some kind, although not everyone might agree what these are. For a person who values non-discrimination above safety, facial recognition systems that discriminate have disvalue embedded in them. For someone who values safety above non-discrimination they have value embedded in them. But they cannot be completely neutral, or they would not be put to use in the first place[13].

To the extent that digital ICTs embed these emergent values, can they then be said to embody a non-human agency that produces outcomes of value/disvalue? This idea requires both a consideration of agency in general, and moral agency in particular. While in the broadest sense, anything that causes an outcome can be said to be an agent (Schlosser, 2015), we generally construe agency as the the exercising of a capacity to act. Given that digital ICTs can both act and produce actions, are they arguably potentially artificial agents?

Grodzinsky, Miller and Wolf (2008) suggest that an artificial agent is "a nonhuman entity that is autonomous, interacts with its environment and adapts itself as a function of its internal state and its interaction with the environment" (p. 116). This definition seems to apply to many digital ICTs, but does not address the question of intentionality, which on the standard definition, is required for agency. For the purposes of this argument, I am going to assume however, for now, that digital ICTs are indeed potentially some kind of artificial agent, holding off a discussion on intentionality until a later point.

Assuming then, that digital ICTs are agents, the question follows as to whether outcomes of moral value/disvalue necessarily imply that the digital agent is a moral agent.

---

[13] Furthermore, because these values are emergent *and* embedded, they are also to a certain extent dependent on the particular context of use of the technology. This may seem like a relativist position, but in fact what is mutable is the technology/context/use actor-network, not the assessment of value/disvalue, and so I would suggest that this is not a relativist argument to make. That context is crucial is important, however, and will be discussed in detail later in this essay, with reference to Nissembaum's contextual spheres.

Moral Agency

Agency can be generically understood as the capacity to act (Himma,2009), but it is immediately obvious that not all agency is necessarily morally significant. My cat acts when it catches a mouse, but I would not necessarily want to attach moral significance to that, although I might feel sorry for the mouse and there are clearly consequences to the action (being of value to the cat and disvalue for the mouse). What makes agency morally significant? How can we define "moral agency"?

As Himma (2009) argues, the "standard account" of criteria for moral agency suggests that deliberation and free will, and therefore consciousness, are necessary and sufficient conditions for an agent to be held morally accountable for her behaviour.

Himma (2009, p.19) begins by defining an agent as someone/something which can do something or cause a performance (i.e. act) in a purposeful or intentional manner. Some kinds of doing are not intentional, even for humans - for example: breathing, growing or digesting food. While these actions have the broader purpose of keeping us alive they are not actions which we directly initiate or cause. We cannot decide *not* to breathe (not for very long anyway), grow or digest food once it is eaten, any more than we decide to do these things.

Action or acting, in contrast, is a special kind of doing, one that has an intention - for example: typing, walking, drinking coffee. To do something purposefully requires an intentional mental state - I have to think about typing in order to type[14], or decide to have a cup of coffee in order to drink it. Therefore, Himma (2009, p.20) concludes, "only beings capable of intentional states can be an agent." Even my cat must think about the mouse, at some cat-level of thinking, in order to catch it.

Moral agency however, means that the agent is governed by moral standards and therefore has moral obligations. A moral agent is accountable for whether her actions

[14] A touch typist of course does not actually think about the individual actions of typing, but the overall action of typing a document (for example a research report), is definitely one that requires purpose and intention.

meet or don't meet those moral obligations. Himma (2009, p.22) suggests that to hold an agent accountable is to give her what is deserved - praise, blame, punishment, reward - as a consequence of her actions. A moral agent, therefore, is one that can rationally be praised, blamed, punished or rewarded for her actions. It is not rational to blame a cat for catching a mouse[15].

According to Himma (2009) there are two conditions required for such desert to be rational. Firstly, it is not rational to hold an agent responsible for an action unless she has freely chosen that action. If I, for example, steal a car because someone threatened my life unless I did so, I am surely not blameworthy for the theft? The person who threatened me is the one who deserves the blame and punishment[16]. Free will, therefore, is a necessary condition for moral agency. Since free will requires choice, and choice requires deliberation, the capacity for deliberation is also a necessary condition for moral agency.

Secondly, it is also not rational to hold an agent morally accountable for her action if she did not understand the rightness or wrongness of that action when she (freely) chose it. Therefore the capacity for moral reasoning is a necessary condition for moral agency. Moral reasoning means that the agent understands moral concepts and basic moral principles and can apply these to the specific case.

Free will, deliberation and moral reasoning all require not only that the agent has an intentional mental state, but also that the agent has access to that mental state. To freely choose to do something means that I am aware that I am doing it. I have to hold in my mind the conscious awareness of the choice, as well as the content of the options. To be aware of the moral import of the choice requires that I consider whether I think the choice to be a good or bad one, and why. So unlike my cat, who

---

[15] I might scold the cat in the hope that this will deter her from catching mice in future. This is however a behaviour modification exercise, rather than a question of just desert. It is also not in my experience very successful.

[16] This is, of course, not an uncontentious position. As argued by Harry Frankfurt (1969 "Alternative possibilities and moral responsibility", *Journal of Philosophy,* 66(23): 829-39), one can sometimes be morally responsible for an action even though one could not have done otherwise, if, had one been free, one would have chosen it anyway.

presumably sees a mouse, thinks "mouse means 'catch it!' ", and does so (whether hungry or not), a morally accountable cat might see a mouse, and then think: "I have a choice whether or not to catch this mouse. The consequence of that will be bad for the mouse, and neither good nor bad for me, since I am not hungry. Since the overall consequence is negative, catching the mouse is a bad choice." Therefore, the morally-accountable not-hungry cat will not catch the mouse. Unfortunately for mice, it seems cats are not capable of accessing their mental states in this way.

Mental states of which we are aware are, according to Himma (2009 p.27) "conscious mental states; one cannot introspect or observe what is not available to consciousness". On this reading "consciousness" is "state consciousness", as opposed to sentience, wakefulness, self-consciousness or transitive consciousness (awareness of other things) (Van Gulick 2014 pp.7:9). Specifically this state consciousness is a state one is aware of being in - that is, a mental state about a mental state. If moral agency requires us to be aware of our actions, as well as about our choices about our actions, then Himma (2009, p.28) concludes, moral agency requires (state) consciousness.

Following this conclusion, it seems evident that a digital agent can be a moral agent only if it is conscious. Although ICTs may be sufficiently sophisticated so as on occasion to *appear* conscious, there is as yet no digital agent that we know of that is to our knowledge state conscious - i.e. aware of its own mental state. Even if we remain, as Himma (2008 p.28) is, "agnostic" as to whether that is possible[17], that does leave us currently in a position where it doesn't seem rational, on the standard account, to hold digital agents morally accountable.

If digital agents are *not* moral agents - not morally accountable - then whom are we to hold accountable for the morally significant consequences of digital agents' actions? Johnson and Miller (2008, p.131) suggest that when it comes to questions of responsibility and accountability for moral consequences, the important thing is keep

[17] Dreyfus (1972, *What computers can't do: A critique of artificial reason*., New York: HarperCollins) provides a compelling phenomenological argument as to why conscious computers are in fact not possible. However, for the purpose of my argument, it is not necessary for that to be the case.

technology “tethered” to the designers - to keep holding the designers accountable. However, this is surely not practical or even possible when values are emergent. If users have flexibly interpreted a design, such that the value in use no longer reflects the intentions of the designer, it may not make sense to hold the designer responsible for that use. If a thief interprets the design of an ATM to enable him to steal from a hapless user, is the ATM designer responsible? Do we blame the victim for being careless? Or is the thief accountable? If the “thief” is in fact an autonomous digital agent, and *not* morally accountable does that mean no-thing is responsible or accountable? That would mean that there is no possibility of applying any moral principles to some digital privacy problems, or even making ad hoc moral judgements about them. We would, in many cases, simply have to shrug our shoulders and admit that digital privacy problems are beyond moral reasoning.

If, on the other hand and as I have argued above, it makes sense to consider, at the object-oriented LoA, the emergent values of both digital and human agents – the values that are being realized *through* action – then it might make sense to consider what the actions are *about* – their intentionality?

Intentionality

“Having content, being about something” is what we call intentionality (Cole 2014).

Since digital systems certainly have content we might ascribe intentionality to them. In that case we would also have to ascribe intentionality to works of art, books, letters, photographs or any representational object at all. In most standard definitions, however, intentionality is assumed to be a mental state (Jacob, 2014), and therefore only a conscious mind is capable of intentionality. Paintings, books, letters and photographs do not, after all, know what they are for. One of the most famous arguments for the necessity of self-consciousness for understanding, intelligence and intentionality is John Searle’s Chinese Room Argument (CRA), best summarized by Searle (1999) himself:

> Imagine a native English speaker who knows no Chinese locked in

> a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese. (p.115)

There are many supporting discussions and rebuttals of the CRA, most of which are directed towards cognitive science and the possibility of artificial intelligence. What is important for this argument is the question of intentionality, and whether a digital agent can ever be said to have intentionality.

If Searle qua the man in the Chinese room is an analogue equivalent of a digital agent – initiating outcomes of value/disvalue, but no consciousness or understanding of the broader context – can he be said to have intentionality? Originally, in the 1980 paper, Searle used the argument to refute the possibility of computer *understanding* (Searle, 1980). Later he extended the implication to refute intentionality as well, writing in 2010 that "… the implementation of the computer program is not by itself sufficient for consciousness or *intentionality*" (Searle 2010 p.17, my emphasis). But surely the messages have content?

As Searle (1980) originally points out, in the case of the Chinese room, the point about the Chinese writing is precisely that it has, in the context of the thought experiment, no content whatsoever. It is simply a formal system of signs –"Squibbles and Squobbles" - with no meaning at all for the person locked in the room. "Because the formal symbol manipulations by themselves don't have any intentionality; they are quite meaningless; they aren't even symbol manipulations, since the symbols don't symbolize anything." (p.11) Searle has designed the thought experiment in this way, because the purpose of the experiment is to refute the possibility of understanding on

the part of computer programmes, which themselves don't understand their own symbols. As he points out (p.13) , "… if you type into the computer '2 plus 2 equals?' it will type out '-4.' But it has no idea that -4" means 4 or that it means anything at all."

Even where symbols or language do have content, this content or intentionality is, according to Searle, indirect rather than original or intrinsic. Language of any kind has content only insofar as it is interpreted by someone who necessarily has a conscious mental state, "caused by and realized by the structure in the brain".

Nonetheless, the *whole project* of translating Chinese must be *about* something. Arguably we could suggest that the entire Chinese Room thought experiment (rather than the person inside the room, or even the whole Room) does have intentionality, because it is *about* demonstrating the impossibility of understanding on the part of an individual programme or algorithm of symbol manipulation. In other words the construction of the system or programme as a whole has a purpose, a reason for being in the first instance. This is not the same as saying "it is the whole room, the whole system that understands Chinese, not the man", which is, as Searle (1999, p. 115) points out, one of the standard objections to the thought experiment. Rather I am suggesting that the fact that the Chinese Room experiment has a particular outcome (demonstrating that the "machine" - the Room - doesn't "think") means that there is content, intent, reason to be (even if only in an imaginary state) attached to the Chinese Room thought experiment *as a whole*. Searle would presumably rebut this point [18] by saying that it is he Searle (qua designer of the system) who has intentionality and not the system itself.

In the case of the Chinese Room, which is a very simple system, it does indeed seem logical to argue that the intentionality is entirely Searle's intentionality. But ICT systems are not simply input-process-output systems like the Chinese Room. Rather they are complex socio-technical systems. As I have argued above, we can seldom attribute all content and "aboutness" in a fully functioning socio-technical system

[18] He rebuts the idea that the whole room understands Chinese in Searle (1999, p. 115), but that isn't the point I am making here.

simply to the intentions of the designer. Interpretive flexibility means that these systems as a whole will come to have emergent embedded values – or things that they are about and for. Can we not therefore ascribe some sort of intentionality to these kinds of systems?

According to Daniel Dennett (2009) this would be a perfectly legitimate thing to do. Dennett suggests that we are justified in adopting an intentional stance towards any complex object whose behaviour is adequately predicted by assuming that it does in fact have intentionality. According to Dennett all intentionality is derived as far as we can know – in other words it doesn't make a difference whether an agent has direct or indirect intentionality, if they are observed to act with what appears to be intentionality. Intentional systems theory "is a theory about how and why we are able to make sense of the behaviours of so many complicated things by considering them as agents" (Dennett 2009, p.349).

According to this theory, the way in which we cope in the world is to *assume* intentionality on the part of many complex objects, such as animals, systems, human beings, cells, eco-systems and so on, because that is how we can, for the most part and with reasonable or good-enough accuracy, get on with dealing with them. We don't need to do this with objects that have static content (books, works of art, etc[19]), but as soon as an object exhibits behaviour, this is how we manage our relationship to and dealings with it. On this basis, any object that produces (rather than merely has) content can be said to be intentional. The object doesn't require a mind, let alone a conscious one in order for us to *infer* intentionality.

Would it then be valid to suggest that intentional systems theory is necessarily at the same LoA as an object-oriented theory of the nature I'm suggesting? In fact it is not. Dennett is quite committed to all intentionality being derived *as far as we know*, at any given LoA. Specifically he claims that

> "(1) there is no principled (theoretically motivated) way to distinguish 'original' intentionality from 'derived' intentionality,

[19] Although of course any critical engagement with a work of art or book will have to ask what it is about, rather than simply what it contains.

and

(2) there is a continuum of cases of legitimate attributions, with no theoretically motivated threshold distinguishing the 'literal' from the 'metaphorical' or merely 'as if' cases." (2009, p.342)

Of course, like the Chinese Room, the perception of intentionality, whether derived or original, requires an agent that is itself capable of making sense of the behaviour *as intentional* in the first instance - that is, that agent must have original intentionality itself. A world in which there was only derived intentionality would in fact be a world without intentionality: A world full of Chinese Rooms - with no John Searle or any one else to argue with him or each other about meaning - would be quite meaningless. As a further example, if humanity were entirely wiped out, but computers survived, some of these computers might be chess computers set up to play chess at the moment of Armageddon. They would then continue to play matches against each other into infinity (or for at least as long as the electricity supply lasted). In such a world, is anyone still 'playing chess'? As per the Chinese Room Argument, a chess playing computer isn't really playing chess, it is executing a particular program, that's all.[20] The same applies for a world full of 'derived intentionality'. It arguably applies today, at the purely technical level of Gmail screening my emails, or Facebook storing pages of Max Schrem's data that may never be used.

This is a very strong argument, in my view, for intentional systems theory not necessarily being applicable at all levels of activity - in other words, it is, per the argument above, difficult to attribute any intentionality at all, if the attribution is not itself being made by an agent that itself has some degree of original intentionality. However, for the specific purpose of developing an object-oriented privacy ethics, if we can adopt a view (a Level of Abstraction) at which we deal with and manage our relationships to other agents (both human and digital) by *assuming* intentionality on their part, and such a Level of Abstraction enables us to make sense of the content and value of their actions, then we may find it useful to do this.

[20] My thanks to Dr Kowalenko for this counter argument and example.

This raises the question of exactly what is meant by a "Level of Abstraction", and what the object-oriented Level of Abstraction that I am proposing would entail.

An Object-oriented Level of Abstraction.

In the discussion above I have argued that IE does not provide a coherent theoretical basis (a macroethics) for considering problems of digital privacy. Where I do think IE is useful is in that Floridi introduces the concept of a Level of Abstraction (LoA) as a conceptual device which can be applied to moral problems, and in particular moral problems involving digital agents.

Floridi initially (1999) suggests that IE is a 'perspective" (p.49) or a way of seeing the world. In this perspective, everything is to be viewed at the level of its information - "An entity is a consistent pack of information…and can be named or denoted in an information process….IE treats every logically possible entity as an information entity" (p.44). An agent is then any entity that can "affect the infosphere" - even by means of its existence. From this perspective, moral judgements are made in terms of how entities affect the infosphere - and it is this which makes IE "object-oriented" (p. 49)- it has all objects in view, both agents and patients, of all kinds, thus "enlarging the ethical discourse" (p.50).

In later writings Floridi (2002, 2008c) formalises this notion of a "perspective" to that of a LoA, a concept which he imports from computer science, where it is used to model informational systems. A level of abstraction is exactly that - a particular level at which we abstract certain details for the purposes of representing reality. So in a computer system a real-life studying breathing existing human being may be represented, for example, by a student number, with which a particular set of courses, course marks, results and an outcome would be associated. Floridi (2002) points out that a LoA provides a "set of observables available at that level" - "the higher the LoA, the more impoverished is the set of observables, and the more extended is the scope of the analysis" (p.288).

Given that we always have a particular perspective when considering a phenomenon

or set of phenomena, or a moral problem for that matter, we can be said to have always adopted some or other level of abstraction. In other words a LoA is "implicit" - the "hidden parameter" or "context" that "allows for a proper definition" - "whether it be in the realm of Euclidean geometry, quantum physics or commonsensical perception" (Floridi and Sanders, 2004, p.353). For Floridi (and computer and other scientists in general) any situation is also "a collection of observables, each which has a well-defined possible set of values or outcomes" (p.354). This leads him to conclude that the Method of Abstraction consists of formalising the model through the analysis of the system (p.355), and that therefore the concept of the LoA is an "epistemological levelism" (2008c, p.304). In other words, the concept of a Level of Abstraction is a device for understanding what is going on, in a particular situation, with a particular epistemological purpose in mind. In order to make sense of - or make an analysis of or moral judgements about - a situation, we have to abstract the relevant features of the situation and focus on those. A Level of Abstraction simply denotes the boundary of the situation and the granularity with which we carve out these features, and therefore what counts in the analysis.

Lucas (2012) points out that, because systems are constructed (rather than naturally occurring), their boundaries are too, and there may therefore be no 'natural' LoA at which we can naturally accept that systems are moral agents (p.49). In fact, "the idea of multiple (and related in some strong sense) LoAs also leads us to ask whether some different LoAs might simply be a case of Wittgenstein's seeing as" (p.50).

Lucas makes this critique of Floridi to argue against an LoA as a "natural characterisation of morality" (Lucas, 2012, p.63). However, for my purposes, "seeing as" is precisely what I am aiming at - an epistemological level or LoA at which we *see* human and non-human agents *as* intentional moral agents, because this is a level at which we are able to make sense of the moral import of their interaction. This is after all the purpose of abstraction: When I equate a dot on a map to a city, it is not that I expect that the dot corresponds in any real way to the city - the city is not a big dot, it is not dot-shaped, it is not a solid entity, is not actually fixed, it is mutable and developing. Nonetheless for the purpose of orienting my travel towards the city, the dot on the map is a truthful representation that allows me to make sense of my

direction. Similarly, an epistemological LoA, at which all agents are seen as intentional, can assist us in making moral judgements about what those agents, in any given situation, ought to do.

What I am *not* arguing for is IE itself - an infocentric LoA at which everything is conceived of as its information - because as shown earlier that doesn't help us with privacy issues. What I am suggesting is a LoA that is

- not anthropocentric (like deontology or consequentialism) or bio-centric (like environmental ethics) but being- or onto-centric - giving moral import to the agency of human and non-human agents alike.
- neither agent oriented nor patient oriented, but object oriented - not only concerned with agents (those acting) or patients (those being acted upon), but with the intersection of both.
- "not an ethics of virtue, happiness or duty, but of respect and care" (Floridi, 1999, p.49). Since I am not arguing for personhood or consciousness for digital agents I cannot frame ethics in terms of virtue, happiness or duty. However, if i am concerned with all agents and patients in the situation I can suggest that we should respect and care for them.

To precisely define the LoA: I am suggesting that in situations of digital privacy concerns, we see that human and non-human (artificial) agents are equally objects producing outcomes of value / disvalue and as such they are considered to have intentionality. Because human and non-human agents take actions that have moral import, they must for purposes of making sense of the world be considered to have moral agency. That is, we can legitimately worry about and judge the moral import of their actions and declare some right and others wrong. The moral imperative is to balance respect and care for both human and non-human agents.

This LoA is of course pragmatic, rather than essentialist. I am not suggesting that non-human agents necessarily and always are intentional moral agents. I am following Dennett (2009) in attributing intentionality to all agents, but only as far as the LoA requires, and in fact not as far as Dennett would go, because I think the counter-argument for requiring conscious intentionality at some level is very strong. So I am

not suggesting that this LoA is essentially and always the only correct or relevant LoA, but rather that it can be useful to apply it *in situations of digital privacy concern,* because this LoA is the workable one, at which we can in fact even begin to make sense of these concerns.

This LoA is usefully applied because at the moment we assume very little moral import attaching to non-human action, and therefore we possibly neglect care and respect for human agents and/or patients in the system.

What follows further, is that at this level human and non-human agents legitimately have competing interests. This raises the question of whether there is a hierarchy of priority of those interests, or at least a principled way to balance them. Given that my intention here is to consider, in particular, the privacy problems posed by digital ICT, it is inadequate simply to say that human and artificial agents are both agents and patients in privacy-related moral action, whose competing rights and interests need to be balanced. If we consider the practical case of Max Schrems and Facebook, it is not adequate to say that both have interests in the case - although it is easy enough to identify them - Max Schrems wants privacy and his data to be deleted, and Facebook wants to keep it (and all the other data of the billions of people on Facebook, which is the point of the legal case of course). We have to ask what intrinsic values underpin those interests, and how we can sensibly consider them. We need to consider on what basis we can evaluate and balance interests in respect of digital privacy in order to apply the object-oriented LoA to digital privacy problems.

Applying an object-oriented LoA to digital privacy problems

As discussed earlier in this essay, it is precisely the difficulties of "privacy in public" caused by the use of ICT to reveal identity in very profound ways (Nissembaum, 1997, 1998) that we need to deal with. If it is in the interests of the non-human agent to whom I have freely given my data that it is used, then what right do I have to preclude its use? If Max Schrems chooses to share his information with the world using Facebook, for free, then why shouldn't Facebook itself store, mine and sell that information? Isn't that a fair exchange? What information is private and what is

simply public anyway?

If moral constraints about the use of personal data exist only to prevent against intrusion into the private realm, whether by governments, corporations or other individuals (Nissembaum, 2004), then data that is already in the public realm - whether by virtue of its location or its nature - cannot be said to be protectable in the first place. This is, however, at odds with both our intuitions and the demonstrable potential for intrusion inherent in the use of publicly-available data by powerful ICTs.

As noted by Nissembaum (1997, 1998, 2004), Schonscheck (1997) and Van Den Hoven (1997), the reason that privacy in public places is not well accounted for by traditional theories is that these theories assume a dichotomy between the public and the private. This is an over-simplification of the complexities of what data (with respect to both type and location) can be considered to be private. Following Schoeman (1984, cited in Nissembaum, 2004), these authors variously note that there are in fact multiple private realms or spheres. Privacy is therefore something that pertains within a particular sphere with respect to norms of appropriateness (Nissembaum, 2004, p.138) and distribution (Nissembaum 2004, p.140).

Appropriateness is a norm that dictates what kind of information it is appropriate to reveal in a particular context or sphere. The relevant criteria for judging whether a piece of data or information should be shared is thus associational relevance, rather than whether the information itself is public or private, or of a sensitive nature. Familiar examples include the extent to which I would share medical information with my doctor, or financial information with my bank.

Once we perceive that the desired privacy of particular kinds of data is relative to particular contexts or spheres, then it becomes evident that privacy concerns - and particularly concerns of privacy in public - are raised by the distribution of data within and even more so between spheres. In fact, in many cases, we would not want data to be shared between spheres at all - we don't want our medical records given to our bank or our financial information to our doctor. Publicly-available data that is mined to provide a rich "mosaic" (Schonsheck, 1997, p.225) of our patterns of behaviour -

and then sold to corporations that may be well outside the sphere in which the data was originally collected or provided - is very likely to be data that has not only crossed multiple spheres, but has changed in its associational relevance along the way.

Both Nissenbaum (1997,1998, 2004) and Van Den Hoven (1997, 1999) follow Michael Walzer's (1983) idea of spheres of social justice, within which particular norms of distribution apply, and which should be impermeable. Walzer has in mind social goods such as power or money, which may legitimately be distributed in particular ways within a sphere (money in the commercial sphere, power in the political sphere) but which should not have influence across spheres. So money should not be able to buy political influence, for example. Nissenbaum and Van Den Hoven (separately) suggest that information is equally a good that, given particular norms of association, should not be distributed across discrete spheres. As Van Den Hoven puts it "what is often seen as a violation of privacy has more to do with information traffic across the boundaries of what we intuitively think of a separate social spheres" (1999, p.144)[21].

This concept is useful for understanding why it is that we intuitively feel that privacy has been breached when information is disclosed in unexpected ways or places. It does however have its limitations, if it is bound to social norms (Nissenbaum 2004, p. 144) of what information pertains to which spheres, rather than moral rules per se. We therefore need to try and identify how norms of distribution can be established that are not simply practice or convention, but rather tied to some intrinsic rights or values. In other words, we can suggest that data should be distributable, or not, between spheres based on an evaluation of the effect that the distribution would have on the intrinsic values of equality, freedom, autonomy or safety, with a view to balancing those across agent and patient, as well as digital and human agents and patients, in order to promote their respect and care.

---

[21] Nagenborg, M. (2009) Designing spheres of informational justice, *Ethics and information technology*, *11*(3), pp.175-179 critiques Van Den Hoven's and Nissenbaum's use of the concept of spheres, on the grounds that if there is a "sphere of information" then sharing, for example, medical information from the economic sphere would not be an injustice, as the "information accessed stills remains within the sphere of information". This seems to be a misreading of both Van Den Hoven and Nissenbaum, neither of whom suggest a sphere of information at all, but rather that information is one of those *goods* (like money, power or education) that should be appropriately distributed within and not between spheres.

How can I justify equality, freedom, autonomy or safety as intrinsic values? A discussion of intrinsic versus extrinsic value is beyond the scope of this essay, except to say that "That which is intrinsically good is nonderivatively good; it is good for its *own* sake." (Zimmerman, 2015 p.6). This particular list of intrinsic values is derived from Nissembaum's (1998) discussion of the 'genuine privacy interest' that follows from the need to protect these individual and social values. (pp. 591:593), and also from the fact that - with the exception of equality - they correspond to each of the types of privacy, that is, freedom to decisional privacy, autonomy to mental privacy and safety to physical privacy.

What I am arguing for, therefore, is that there are in fact intrinsic values, and that these are the meta-ethical basis[22] on which the object-oriented approach to digital privacy problems rests. This is of course, slightly problematic in the absence of the autonomous agent, because rights are normally seen as the consequence of a deontological anthropocentric approach. So how do I account for them? It seems to me that a process of reflective equilibrium allows for this: Intuitively we understand that there is something that needs to be protected in digital privacy situations, and that we need to proceed with caution in ensuring this protection. Since neither purely consequential, nor ownership-based approaches help us to do this, we need an alternative explanation for our moral intuition in this domain. The best explanation may well be that there *are* rights and values at risk, and that these potentially attach to all agents in the situation.

This is why it has been important to establish the object-oriented LoA - because we need to be able to attach rights and values to digital agents as well as human ones, if we are to determine normative guidelines for the distribution of information within and between spheres.

---

[22] This is a further distinction between IE and the applied object-oriented approach that I am attempting to develop here. IE is a meta-ethics, in terms of which the basic principle of good is existence of information, and the basic principle of evil is entropy. I am *not* attempting to develop a meta-ethics, but rather an applied approach to a particular domain, which rests on the concept of intrinsic rights as a metaphysical foundation to ground moral judgements. To the extent that 'rights and values' can be seen to be pluralist, this approach would be equally so.

Just as it is not useful to draw a simple distinction between public and private information, it is not useful to draw a simple the distinction between the online and real worlds because data, and the value accruing to data, does leak across these as a question of fact. This is why object-oriented view is useful, because it flattens "digital" versus "real" distinctions. For example, on the object-oriented view, Google scanning my email is *exactly* the same as someone reading it. That accounts for questions that may be raised about privacy implications of Google scanning my email. The question then to ask is – in what sphere is my email and what norms should apply?

This raises a difficulty, because given the communicative and broad-reaching nature of most digital ICTs, it is sometimes hard to determine exactly what sphere a particular piece of information actually belongs to. A useful way to draw the distinctions may be to draw on the intentionality of the agent and patient in a particular sphere - in other words to ask, what is this information used for, in this sphere? A non-digital example would be my communication with my doctor - I am using medical information to inform my doctor so that she can medically assist me, and she is using it for the same purpose, and perhaps also to inform her own practice and experience, maybe even to write it up in a case study in a journal. The extent to which the information can be shared, and what information is shared, will require it to be used for only those purposes, and in a way that does not infringe on my freedom, autonomy, security or equality. So my doctor can share my details, including my identity, with a specialist who will further assist me; but will make sure I am anonymous if she writes up the case in a journal.

In the digital realm we could consider the example of free-to-use email, such as Gmail. In that case I could say that email is used for communication by me, but for revenue generation for Google. On such a consideration, my communication is legitimately distributable within the sphere of revenue generating activity for Google. This does correspond to the practice of how the sphere works in fact - right now Google does draw the boundaries to include revenue generation. For example, if I send emails with the subject heading "text books" on Gmail, I might shortly receive an email from an online store offering me a discount on text books. Clearly Google has sold the data

that I am interested in text books to the online store. The intrinsic value being protected is Google's freedom to operate profitably. At risk are my autonomy to buy books without being intruded upon and my equality since my relationship with Google is a highly unequal one, in which they did not consider it necessary to ask if they could sell this piece of information.

In fact, this has implications for who gets to draw the boundaries of the sphere. Often it is not explicit to the user where the boundaries are, unless they are paying very careful attention. Perhaps an important principle is that the boundaries need to be explicitly negotiated between the agents – I get to say what I intend to use my email for and what I expect the norms to be, Google gets to say what it intends to use it for – and if we don't agree on that, then we don't enter the sphere at all.

On balance since the digital-objects are generally much more powerful than the human ones (given data volume and processing capacity), this must mean far more disclosure and negotiating power should be given to the human objects than is currently the case. To put it simply, instead of burying privacy policies on pages that need to be specifically accessed, and couching them in lengthy and sometime legalistic terms, organisations need to be really explicit, in simple lay-persons terms, about what they are going to do with my data. Ideally they should point out every time they are using my data. Equally, I need to accept that Google does have some claim, in terms of the sphere in which we co-operate, to make use of that data.

Therefore, I suggest that in order to make sense of digital privacy problems we need to apply the following principles:

1. All agents and patients in the system need to be taken into account
2. The contextual sphere within which the information concerned is shared needs to be defined. This can be done by determining what the purpose of the informational sharing is, in the particular sphere.
3. Movement of information within and particularly beyond the boundary of that sphere is only permissible if it does not compromise the freedom, autonomy, security or equality of either agent or patient.

Would an application of these principles enable us to consistently and sensibly assess our moral intuitions about what is morally permissible and impermissible when it comes to breaches of our privacy online? In other words, are the problems that occur with purely deontological and consequentialist approaches avoided with an object-oriented ethics of this kind?

Certainly any conceptual limitation on applying moral norms to non-human and even non-corporeal (digital) agents is overcome by moving to the level of object orientation I am proposing. The questions to ask are whether this is intellectually defensible, and whether it is useful.

If moving to an object-oriented LoA as I have described it were simply a mental act of imagination or fiction then it would not be intellectually defensible. It would be as if I said, "imagine there are bad fairies in the world" and suggested that we should adjust our behaviour accordingly. Perhaps we would then avoid ever touching a needle, but this would be simply superstition. As I have demonstrated however, digital systems do have emergent values that are independent of their designers, and which we need to take into account. So conceptually viewing systems as having moral intentionality is not just a fiction but rather a particular view or abstraction. It is a legitimate way of knowing about these systems and our interactions with them that is not in the realm of the merely fictional.

The question remains as to whether the approach is useful. If we apply it to a particular case we may be able to see whether it helps. Let us take the case of Max Schrems and the thousands of pages of his data that Facebook stored without his knowledge. Applying the principles I have listed above:

1. Both Schrems and the Facebook system are valid objects of consideration in this case;
2. The contextual sphere is one in which Schrems voluntarily shared information about himself, and presumably his personal life, with anyone who had access to his Facebook page, and therefore with Facebook itself. For Schrems the purpose of this sphere is to share information, for a limited period of time (since he deleted it), with his friends. For Facebook the purpose of collecting Schrems information

in the first place, and then storing it, is to be able to run a profitable organisation with this kind of data.

3. The movement of data up to now appears to be within the sphere. Schrems 'deleted' it, which in practice meant that Facebook 'moved' it from being accessible to Schrems and his friends, to being stored - or accessible to Facebook only (we can for the purposes of this case assume they didn't sell it as such). Did this movement of data compromise the freedom, autonomy, security or equality of either Facebook or Schrems? Arguably this movement, while promoting Facebook's freedom (to do with the data as it saw fit) and security (as a viable commercial concern), compromised Schrems' autonomy (his knowledge of and ability to change something that relates very closely to him) and equality (knowing as well as Facebook does what is going on), precisely because he did not initially know that Facebook either would or had stored this data.

I would suggest that the object-oriented approach *is* useful, because it shows up some important aspects of the case. Firstly it highlights that Facebook does have some interest that needs to be taken into account, and it is not adequate simply to cast it as the Big Bad Wolf in the story. Secondly, it explains why someone like Schrems intuits that what Facebook has done is not acceptable - it is because Facebook kept his data, secretly, without telling him what it was doing. In this way his autonomy and his equality were both compromised. Since Facebook's action in keeping the data compromised these rights, it is morally impermissible. Were Facebook to be completely transparent with all users about what it does with 'deleted' data - and the users consented by using Facebook anyway, it would be morally permissible.

Conclusion

I can, therefore, conclude that the object-oriented LoA that I am proposing can be adopted for specific cases - that is, for digital privacy problems. In such cases, and for the specific purpose of weighing up the competing rights and values of all the agents and patients in the case, we can treat all agents (human and non-human) as both intentional and real. This provides a reading of the case that goes beyond the consequentialist or ownership-based approaches, and arguably gets closer to the

heart of the issue.

Where the approach is still open, however, is that we still somehow have to justify and balance these interests. We have to find a way to respect and care for both digital and non-digital agents and patients. So while I have proposed some principles for balancing interests, based on the concept of context spheres, there are still issues remaining in how we approach digital privacy problems.

The first issue is to justify more carefully the selection of rights and values that I have chosen to specify as intrinsic. I have attempted a parsimonious and fairly obvious list, but perhaps some of my rights are cultural not intrinsic. There also are probably others that I should have included. The second issue is to consider further how we balance interests once we have specified them. In the discussion above I have suggested that if any of the intrinsic rights are violated, then the action is morally impermissible. The Max Schrems case seemed to suit that inference. There are however other instances in which we might want to violate a right because another right is more important. For example would we hesitate to violate the online privacy right of a terrorist using Facebook to plot murder?[23] This is a simple example, but there are undoubtedly others in between these two examples, that are more complex to consider.

So while the framework of the contextual sphere provides some additional reasoning for why particular interests are valid in any particular case, I would suggest that there is probably no simple formula to apply. This doesn't seem counter-intuitive or incorrect to me. If moral problems were solvable by equations they wouldn't really be moral problems at all, they would be mathematical problems. So, in the end, it not surprising that a need for practical wisdom, or Phronesis, in the form of a judicious weighing of moral interests continues to apply to digital problems posed by ICT.

---

[23] Even this case is not as simple as it seems. Apple is, as at February 2016, refusing to accede to an FBI request for Apple to amend the operating system of an iPhone, so that the FBI can gain access to the records of the San Bernadino killer, citing data security concerns in Cook, T. (2016) *A message to our customers*, 16 February, URL=<http://www.apple.com/customer-letter/>

References.

Akrich, M. and Latour, B. (1992) “A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies”, in Bijker and Law (eds.) *Shaping Technology/building Society: Studies in Sociotechnical Change*, Cambridge, Mass: MIT Press, pp.259-264.

BBC News (2014), “Facebook privacy challenge attracts 25,000 users”, URL: http://www.bbc.com/news/technology-28677667 last accessed 13 April 2015.

Bonfanti, M. E. (2014) From sniffer dogs to emerging sniffer devices for airport security: an opportunity to rethink privacy implications?, *Science and engineering ethics*, *20*(3), pp.791-807.

Brey, P. (1997) Social Constructivism for Philosophers of Technology: A Shopper's Guide, *Techné: Journal of the Society for Philosophy and Technology,* 2(3-4), pp.56-78.

Brey, P. (2000) Disclosive Computer Ethics, *SIGCAS Computers and Society*, 30(4), pp.10-16.

Brey, P. (2008) Do we have moral duties towards information objects?, *Ethics and Information Technology*, 10, pp. 109-114.

Brey, P. (2010) “Values in technology and disclosive computer ethics”, in Floridi (ed.) *The Cambridge Handbook of Information and Computer Ethics*, Cambridge: Cambridge University Press, pp.41-58.

Capurro, R (2008) On Floridi’s metaphysical foundation of information ecology, *Ethics and Information Technology*, 10, pp.167–173.

Christman, J. (2015) "Autonomy in Moral an Political Philosophy", in Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), URL = <http://plato.stanford.edu/archives/spr2015/entries/autonomy-moral/>.

Cole, D. (2014) "The Chinese Room Argument", in Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), URL = <http://plato.stanford.edu/archives/win2015/entries/chinese-room/>.

Dennett (2009) “Intentional Systems Theory”, in Maclaughlin, Beckermann, Walter (eds.) *The Oxford Handbook of Philosophy of Mind*, Oxford: Oxford University Press, pp.339-350.

Duhigg, Charles (2012) “How Companies Learn Your Secrets”, *New York Times*, 16 February 2012, URL= http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=6&_r=2&hp, last accessed 16 April 2015.

Elton, B. (2007) *Blind Faith*, London: Bantam Press.

Evans, D. (2011) *The Internet of Things: How the Next Evolution of the Internet is Changing Everything*, Cisco Internet Business Solutions Group White Paper, San Jose CA: Cisco Inc.

Floridi, L. (1999) Information Ethics: On the Philosophical Foundations of Computer Ethics, *Ethics and Information Technology*, 1(1), pp.33-52.

Floridi, L. (2002) On the intrinsic value of information objects and the infosphere, *Ethics and Information Technology*, 4(4), pp.287-304.

Floridi, L. (2005) The Ontological Interpretation of Informational Privacy, *Ethics and Information Technology*, 7(4), pp.185-200.

Floridi, L. (2008a) "Foundations of Information Ethics”, in Himma and Tavani (eds), *The Handbook of Information and Computer Ethics*, Hoboken, New Jersey: John Wiley and Sons, pp.3-23.

Floridi, L. (2008b) Information Ethics: A Reappraisal, *Ethics and Information Technology*, 10, pp.189-204.

Floridi, L. (2008c) The method of levels of abstraction, *Minds and machines*, *18*(3), 303-329.

Floridi, L., and Sanders, J. W. (2004) On the morality of artificial agents, *Minds and Machines*, *14*(3), pp.349-379.

Friedman, B., and Nissenbaum, H. (1996) Bias in Computer Systems, *ACM Transactions on Information Systems*, 14(3), pp.330–347.

Grint, K., and Woolgar, S. (1995) On some failures of nerve in constructivist and feminist analyses of technology, *Science, Technology, & Human Values*, 20, pp.286–310.

Grodzinsky, F., Miller, K. and Wolf, M. (2008) The ethics of designing artificial agents , *Ethics and Information Technology* 10, pp.115–121.

Haraway, D. (1991) *Simians, cyborgs and women: The reinvention of nature*. New York: Routledge.

Himma, K. E. (2009) Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent?, *Ethics and Information Technology*, *11*(1), 19-29.

Introna, L.D. (2005) Disclosive ethics and information technology: disclosing facial recognition systems, *Ethics and Information Technology*, 7, pp.75-86.

Ribeiro (2016) "Uber to Pay $20,000 in Settlement on Privacy Issues with New York Attorney General", *IT News*, 6 January, URL=<http://www.itnews.com/article/3019994/uber-to-pay-20000-in-settlement-on-privacy-issues-with-new-york-attorney-general.html>. Last accessed 9 February 2016.

Jacob, P. (2014) "Intentionality", in Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), URL = <http://plato.stanford.edu/archives/win2014/entries/intentionality/>.

Johnson, D. (1994) *Computer Ethics*, 2nd ed., Englewood Cliffs, New Jersey: Prentice Hall,Inc.

Johnson, D. G., and Miller, K. W. (2008) Un-making artificial moral agents, *Ethics and Information Technology*, *10*(2-3), pp.123-133.

Kirkpatrick, M. (2010) "Facebook's Zuckerberg says the age of privacy is over, *ReadWrite*, URL= http://readwrite.com/2010/01/09/facebooks_zuckerberg_says_the_age_of_privacy_is_over, last accessed 16 April 2015.

Latour, B. (1991) "Technology is society made durable", in Law (ed.) , *A sociology of monsters: Essays on power, technology and domination*, London: Routledge, pp. 103–131.

Latour, B. (1992) Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts, in Bijker and Law (eds.) *Shaping Technology/Building Society: Studies in Sociotechnical Change*, Cambridge, Mass: MIT Press, pp.225-258.

Latour, B. (1999) *Pandora's Hope: Essays on the Reality of Science Studies*, Cambridge, Mass: Harvard University Press.

Lucas. L. (2012) "Levels of Abstraction" in Demir, H. (Ed.). *Luciano Floridi's philosophy of technology: Critical reflections, 8,* Dordrecht: Springer Science & Business Media.

Megan Meier Foundation (2015) "*Bullying, Cyberbulling and Suicide Statistics*", URL = http://www.meganmeierfoundation.org/statistics.html, last accessed 16 April 2015.

Messerly, J.G. (2007) Disclosive Computer Ethics?, *SIGCAS Computers and Society*, 37(1), pp.18-21.

Moor, J. H. (1997) Towards a Theory of Privacy in the Information Age, *Computers and Society*, *27*(3), pp.27-32.

Nissenbaum, H. (1997) Toward an approach to privacy in public: challenges of information technology, *Ethics & Behavior*, *7*(3), pp.207-219.

Nissenbaum, H (1998) Protecting Privacy in an Information Age: the Problem of Privacy in Public, *Law and Philosophy*, 17(5-6), pp.559-596.

Nissenbaum, H. (2004) Privacy and Contextual Integrity, *Washington Law Review*, 79(1), pp.119-158.

Orlikowski, W. (1992) The Duality of Technology: Rethinking the concept of technology in organizations, *Organisation Science*, 3(3), pp.398-427.

Rachels, J. (1975) Why privacy is important, *Philosophy & Public Affairs*, pp. 323-333.

Schoeman, F. D. (1984). *Philosophical dimensions of privacy: An anthology*. Cambridge: Cambridge University Press.

Schlosser, M. (2015), "Agency", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), URL = <http://plato.stanford.edu/archives/fall2015/entries/agency/>.

Searle, John. R. (1980) Minds, brains, and programs. *Behavioral and Brain Sciences,* 3(3), pp. 417-457.

Searle (1999), “The Chinese Room”, in Wilson and Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge, MA: MIT Press, pp. 115-116.

Searle (2010) “Why Dualism (and Materialism) Fail to Account for Consciousness”, in Lee (ed) *Questioning Nineteenth Century Assumptions about Knowledge, III: Dualism.* New York: SUNY Press. as cited in Cole (2014).

Schonscheck, J. (1997) Privacy and Discrete “Social Spheres”, *Ethics & Behavior*, *7*(3), pp.221-228.

Schroeder, M. (2012) "Value Theory", in Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition), URL = <http://plato.stanford.edu/archives/sum2012/entries/value-theory/>.

Singer, P. (1972) Famine, affluence, and morality, *Philosophy & Public Affairs*, pp. 229-243.

Sinnott-Armstrong, W. (2014) "Consequentialism", in Zalta (ed.)*, The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), URL = <http://Wplato.stanford.edu/archives/spr2014/entries/consequentialism/>.

Stahl, B. C. (2008) Discourses on information ethics: The claim to universality, *Ethics and Information Technology*, 10, pp.97-108.

Steel, E (2013) For sale: age, gender and location information at $0.0005 per person, *Financial Times*, Thursday June 13, p.1.

Tavani, H.T. (2008) Floridi's ontological theory of informational privacy: Some implications and challenges, *Ethics and Information Technology,* 10*,* pp.155-166.

Van Den Hoven, J. (1997) Privacy and the Varieties of Moral Wrong-doing in an Information Age, *Computers and Society*, 27, pp.33-37.

Van Den Hoven, J. (1999) '"Privacy or Informational Justice?" in Porciau (ed.), *Ethics and Electronic Information in the Twenty-first Century,* West Lafayette Indiana: Purdue University Press, pp.139-150.

Van den Hoven, J. (2008) "Information Technology, Privacy, and the Protection of Data", in van den Hoven and Weckert (eds.) *Information Technology and Moral Philosophy*, Cambridge: Cambridge University Press, pp.301-322.

Van Gulick, R. (2014) "Consciousness", in Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), URL = <http://plato.stanford.edu/archives/spr2014/entries/consciousness/>.

Walzer, M. (1983) *Spheres of Justice,* Oxford: Blackwell.

Zimmerman, M. J. (2015) "Intrinsic vs. Extrinsic Value", in Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), URL = <http://plato.stanford.edu/archives/spr2015/entries/value-intrinsic-extrinsic/>.