UNIVERSITY OF SOUTHAMPTON

A Paradox of Privacy: Unravelling the Reasoning behind Online Location Sharing

by

Aristea Maria Zafeiropoulou

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Faculty of Physical Sciences and Engineering

School of Electronics and Computer Science

November 2014

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

by Aristea Maria Zafeiropoulou

With the emergence of web applications that enabled user-generated content and social interactions, the Web became a place where people can engage in a number of new activities. With the success of smart enabled devices people now actively share their location data through various applications. However, as this thesis reveals, location plays a primal role in linking and inferring new information about people, often without their knowledge or consent. Due to this inferential power of location data new privacy concerns arise, as the actual affordances of their data are far greater than people are even aware of. Regardless of the numerous controversies around privacy, people keep on sharing their data on the Web. However, privacy systems themselves (and the ways in which individuals express their preferences) have changed very little. This thesis argues that understanding the mechanisms that people employ in their privacy decisions can provide fundamental insight for the design of privacy systems. The main focus of this thesis is to understand the underlying reasons why people share their location and whether their disclosure behaviour is paradoxical when compared with their stated attitudes towards location sharing. The first part of this thesis involves a study comprising of an online survey that addresses these two issues. The findings provide supporting evidence that people's location sharing decisions are indeed paradoxical in comparison with their stated attitudes and that privacy decision-making can be seen as a process of structuration, in the sense that people's decisions are tempered by contextual factors (external structures). The second part comprises of a series of focus groups that act as a follow-up study and aim to explore in more detail the underlying reasons behind people's sharing decisions. The findings show that people's decisions are influenced by a number of different contextual factors, grouped together into three main categories; social capital, trust in the application and functionality. Based on the outcomes of the two studies, a conceptual model was developed, called the Isorropic Model, that points out the prominent role of context in privacy decision-making and stresses the need for more dynamic privacy systems.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deep gratitude to my supervisor Dr David Millard for his tremendous support throughout my PhD. With his guidance I was given the opportunity to explore the exciting topic of online privacy.

I am also particularly grateful to my other supervisors Dr Craig Webber and Dr Kieron O'Hara, and of course to the people involved in the Web Science Doctoral Training Centre for offering me the opportunity and the funding to pursue this PhD. I would like to thank the academics and the fellow students of the Web and Internet Science Lab for providing a unique environment that not only promotes creativity, but also offers enormous support.

I also wish to express my appreciation to Professor George Metakides for encouraging me to follow the path of the PhD and Web Science in particular.

Many thanks to Alex Recio for her willingness to offer advice at any time on practical issues of empirical research and for being a great friend.

I would also like to thank Jordi for his continuous support throughout this process that we experienced together.

Finally, I would like to thank my father, my mother and my brother who always encouraged me to pursue new opportunities in my career. Without their support I wouldn't have had the courage to make this step.

This thesis is dedicated to the memory of my father.

Chapter 1

Introduction

> Privacy was, he said, a very valuable thing. Everyone wanted a place where they could be alone occasionally. And when they had such a place, it was only common courtesy in anyone else who knew of it to keep his knowledge to himself.
>
> —George Orwell, Nineteen Eighty Four

As the Web has evolved into an entity where people share their private lives by publishing content online, the boundaries between privacy and publicity have blurred. Privacy protection on the Web is a rather challenging topic. For example, it is common knowledge that Web companies are looking to collect data about their users. Gomez et al. (2009) found that the majority of the most visited websites use personal information for customised advertising, and a large number of technology companies (e.g. Facebook, Google, Yahoo, Microsoft) share their customer data with a large number of their affiliated companies (the average number of subsidiaries was 297, the median was 93). As a result, online privacy has become a subject of heated debate with many groups fighting for the protection of people's data (such as the Electronic Frontier Foundation[1] and Privacy International[2]) and representatives of companies claiming that the end of privacy is here[3].

In the last few years numerous privacy-related issues have appeared in the news, such as the continuous changes in the privacy settings of Facebook or the iPhone location tracking incident[4]; in 2011 researchers discovered that iPhones track all the moves of their users, logging all the information in a single file. A more recent study highlighted people's lack of control over their data by recovering a huge amount of personal data from second-hand Android phones[5]. The list of news stories raising the alarm is endless;

[1] www.eff.org
[2] www.privacyinternational.org
[3] http://www.theguardian.com/technology/2010/jan/11/facebook-privacy/ Privacy no longer a social norm, says Facebook founder, The Guardian
[4] http://petewarden.github.com/iPhoneTracker
[5] http://blog.avast.com/2014/07/08/tens-of-thousands-of-americans-sell-themselves-online-every-day

yet many Web users do not seem to actively think about the long-term effects of their online data sharing and the privacy-related issues that emerge on the Web. According to a survey by Microsoft, less than half of adults and children consider the effects of their online actions on their personal reputation and even less contemplate on the effects these may have in the long run (Brackenbury and Wong, 2013). Still, a report from the Pew Research Centre finds that many Americans faced privacy breaches in 2014. More particularly, 18% of them had important personal information stolen, whereas 21% of them had an email or social networking account compromised[6].

At an individual level privacy is fundamental, as it enables people to maintain their social relationships. People behave differently in different contexts (e.g. at home with their family, or at work with their colleagues). Rachels (1975) provides a detailed account of why privacy is important, where he stresses that different behavioural patterns are associated with different relationships. Depending on the context people may regard certain aspects about them as private, thus a privacy violation may have minor or major implications for them (e.g. losing their job).

As society struggles to understand and manage the privacy issues that have come up with the emergence of digital technologies, the pace of technological innovation continues to increase. The advent of the Social Web and Web 2.0 has caused numerous concerns over people's privacy on a global scale. Although the Web was initially invented with a humanitarian vision — a place where people could meet and interact freely — latest reports reveal that it has been used as a tool for global surveillance. Surveillance mechanisms are often associated with Michel Foucault's panopticon, a concept used metaphorically that is based on Jeremy Bentham's original idea of a prison where prisoners are watched without them knowing when this actually happens. For Foucault modern societies often exercise panoptic mechanisms in an attempt to watch and regulate people (Foucault, 1977).

In 2013 a story regarding online global surveillance broke down turning suspicions of online surveillance into reality and sparking a worldwide controversy. Classified documents from the National Security Agency of the United States (NSA) were leaked to the press revealing that NSA intercepts the communications of people across the globe. The story raises a number of issues regarding surveillance and privacy, the limits of the law, business-government relations, and of course international relations. In essence, it reminds us how important is our need for privacy. The latest news story on this case suggests that photos of people's faces are harvested through various sources (from emails and social media accounts to video chats) at an unprecedented scale to be used in facial recognition programmes[7]. Although users consider these mediums to be private (we debate on this later on in this thesis) and therefore expect a level of privacy, in a

[6] http://www.pewresearch.org/fact-tank/2014/04/14/more-online-americans-say-theyve-experienced-a-personal-data-breach/

[7] http://www.nytimes.com/2014/06/01/us/nsa-collecting-millions-of-faces-from-web-images.html/ NSA Collecting Millions of Faces From Web Images, New York Times

scenario where they did not publish their data online, NSA would not have access to them. Similarly, the success of social network sites and other web applications depends on people's willingness to publish content in these places. With time, though, people are also becoming increasingly dependent on these applications, as online sharing becomes part of their daily routine. In that sense, the question that we raise is to what extent do people consider the long-term consequences of their online sharing decisions and the potential effects to their privacy.

1.1 Focus of this thesis

This thesis aims to explore the underlying reasoning behind people's privacy decision-making. The main hypothesis of this research is the following:

The privacy paradox applies to location data, yet it does not adequately describe the distance between people's beliefs regarding their data and the actual affordances of their data. Privacy disclosure decisions around location can be understood based on discrete factors, including agency and structures, that are mitigated by context.

The *privacy paradox* is a term used to describe the inconsistency that is often observed between people's self-professed privacy attitudes and their actual disclosure behaviours.

The *actual affordances* of people's data refer to the ways systems make use of people's data in practice.

Agency refers to people's capacity to act independently, whereas *structure* refers to the rules and resources that shape people's behaviour.

Context refers to the dynamic social settings under which a privacy disclosure decision takes place.

The hypothesis was broken down into a number of research questions:

1. What are the affordances of location data and to what extent are people aware of them?
   The first question explores the ways that systems manipulate location data and people's level of awareness of these affordances. The motivation behind this is that location is part of someone's physical context, therefore inferences based on location can take place. Thus, it would be of particular interest to analyse in detail the extent to which inferences do take place as well as the extent to which people are aware of these.

2. How do people perceive and value their location privacy in theory?
   This question aims to explore people's attitudes towards the privacy of their location data.

3. How do people value their location data in practice during the privacy trade-off?
   This question focuses on people's actual disclosure decisions and how they evaluate the costs and the benefits of their decisions. The combination of the last two research questions aims to investigate whether people's disclosure behaviours are paradoxical when compared to their privacy attitudes.

4. To what extent do people act as agents and to what extent are they influenced by certain structures during the privacy trade-off?
   This question aims to uncover the extent to which people are constrained not to act entirely freely by certain structures.

5. Can we develop a model of privacy that takes into account the contextual and dynamic nature of location privacy?
   This question was created based on the outcomes of the studies that addressed the previous research questions. The development of a model that places the contextual and dynamic nature of privacy at the centre of its attention can offer valuable support to people's privacy decisions.

6. What are the key factors of this model?
   It is particularly important to identify the key factors of this model based on the actual reasoning mechanisms behind people's privacy decisions.

The first part of this research addresses the first research question. It sheds light upon the various potential contextual elements related to location sharing and brings them together in a study of privacy decision-making. More specifically, it involves the development of a framework for analysis of location data, which is later on employed in the analysis of the data used in a set of technical systems. The analysis highlights the primal role of location data as a starting point for aggregating and inferring other types of data often without the users' knowledge or consent. The analysis is followed by the development of a conceptual model, called the Distance Model of Belief, Behaviour and Affordance, which distinguishes two major disconnects related to privacy decisions. The first stresses people's lack of awareness regarding the ways systems can manipulate their data (e.g. location-based inferences), whereas the second stresses another important issue which is people's paradoxical disclosure behaviour which are often in conflict with their privacy attitudes, a phenomenon known as the *privacy paradox*.

The main part of the thesis aims to answer the following three research questions — i.e. questions 2, 3, and 4 — therefore it explores why people trade their privacy on the Web, a concept known as the *privacy trade-off*, and whether privacy decisions are paradoxical when compared with their stated privacy attitudes. We argue that this

study is important, because understanding the mechanisms that people employ during their privacy decisions can provide fundamental insight for the design of privacy systems — in agreement with Knijnenburg (2013). In this context, the research also investigates the extent to which they act based on their own free will (as free agents) and to what extent they are influenced by certain structures during this trade-off. An example of a structure could be the reputation a system that requests their location has among other users.

An online survey was conducted aiming to address these questions. Hoping to investigate in more depth people's paradoxical behaviours, we developed a set of simple scenarios from people's every day use of web applications. We wished to explore people's responses in each of these scenarios, as well as their justifications behind these responses. The survey was completed by 150 participants, and the results were analysed both quantitatively and qualitatively.

Following the outcomes of the survey, a series of focus groups took place aiming to gain a deeper insight into people's privacy disclosure mechanisms. Three separate focus group sessions took place, followed by a qualitative analysis of the transcribed discussions.

Both analyses (of the survey data, and the data from the focus groups) uncover a plethora of different contextual factors that influence people's decisions. Based on the outcomes of the qualitative analyses of the survey and the focus groups we develop the Isorropic Model, which addresses the remaining two research questions — i.e. questions 5, and 6. The Isorropic Model has important consequences for the understanding of how privacy disclosure decisions are made, and indicates a direction for new privacy systems and interfaces, that emphasise the primacy of context in dynamic disclosure decisions.

1.2 Publications

Parts of the work presented in this thesis have been published as individual research papers. The list below includes a summary of these publications:

- *Privacy Implications of Location and Contextual Data on the Social Web* ACM Web Science Conference 2011 (Zafeiropoulou et al., 2011).
  This paper, presented as a poster paper at ACM WebSci'11, argues for the privacy concerns that are raised from online location sharing that have effects beyond location, since other contextual information can be inferred through location information.

- *Location Data and Privacy: A Framework for Analysis* Reseaux sociaux: Culture politique et ingenierie des reseaux sociaux (Zafeiropoulou et al., 2012).
  This paper, published as a book chapter, presents the framework for analysis of

location data as well as the analysis of a set of technical systems based on the properties of the framework. The results point out the primal role of location information as a starting point for aggregating and inferring other types of information.

- *Unpicking the Privacy Paradox: Can Structuration Theory Help to Explain Location-Based Privacy Decisions?* ACM Web Science Conference 2013 (Zafeiropoulou et al., 2013).
  This paper, presented as a full paper at ACM WebSci'13, describes the study into privacy and location sharing that comprises of an online survey. The quantitative analysis of the data shows the presence of the privacy paradox on location information, whereas the qualitative analysis sheds light on the factors that lie behind it.

Another paper has been submitted and is currently considered by reviewers. This paper, entitled *To Share or Not to Share: The Isorropic Model of Contextual Privacy Decisions*, aims to explain people's privacy decision mechanisms through the development of a conceptual model, called the Isorropic Model. The model is based on the outcomes of the previous study (the online survey) combined with the outcomes of a study that comprises of a series of focus groups. The findings stress the role of social capital, trust in the application, and functionality of the application in privacy decision-making, but also highlight that context plays a primary role in disclosure decisions.

1.3 Document outline

The research presented in this thesis contributes to the field of online privacy decision-making and in particular it explores the underlying reasoning behind people's location sharing on the Web as well as the relationship between people's attitudes towards privacy and their actual sharing behaviour. This thesis is outlined as follows:

Chapter 2 introduces the topic of this thesis by describing background literature on the definition and nature of privacy in the pre-Web but also in the Web era, along with the theories that deal with privacy decision-making that are relevant to this research.

Chapter 3 focuses on privacy decision-making on the Social Web and offers a broad review of current research in this area, including research on social interactions in social network sites, demographic differences, the role of trust in privacy decision-making, and, finally, online privacy management.

The contributions of this thesis are presented in Chapters 4 to 7. Chapter 4 introduces the concept of location privacy and the role of context in location sharing. It continues with the presentation of a framework for analysis of location and contextual data in a set

of technical systems. The chapter concludes with the presentation of a theoretical model of privacy, called the Distance Model of Belief, Behaviour and Affordance (DMBBA). The model points out a disconnect between people's privacy attitudes, their disclosure behaviours, and the actual affordances of their data by online applications.

Chapter 5 explores the relationship between people's attitudes towards privacy and their location sharing behaviours. To that purpose we conduct an online survey that addresses a set of relevant research questions. The full cycle of this study is described in detail, along with the findings of the quantitative and qualitative analysis of the survey data.

Chapter 6 presents the follow-up work of the survey, which consists of a series of focus groups aiming to explore in more depth how people articulate their privacy attitudes and sharing decisions. The chapter presents all the steps we undertake in this study, followed by a detailed description of the findings of the qualitative analysis of the study data.

Chapter 7 aims to bring together the findings of the two main studies. It begins with a discussion of the main findings of these studies (survey and focus groups). With that in mind, we revisit the findings of the two studies and group their themes together. Through the interpretation of their findings we develop a conceptual model that aims to depict the privacy trade-off, which we call the Isorropic Model for privacy decisions. The model hopes to untangle the complex nature of online privacy decisions and inform privacy systems regarding the underlying mechanisms of privacy decision-making.

Finally, Chapter 8 concludes this thesis by offering a summary of the research, outlining its main contributions and proposing potential future work paths.

Chapter 2

Privacy Attitudes and Disclosure Behaviours

Τα εν οικω μη εν δημω.[a]

[a]What happens in one's house should not be broadcast in public.
—Ancient Greek proverb

In the past few years privacy has become a topic of attention and debate in our global society. However, privacy is a rather old concept that roots back to the origins of western civilisation. In ancient Greece, a popular proverb stated that what happens in one's house (*oikos*) should not be made public (*demos*), showing that already at that time there was a clear distinction between private and public. Compared to the laws of the *polis*, one's *oikos* had its own internal rules and was part of one's private life. In modern Greek, the word for privacy is *"idiotikotita"*, stemming from the word *"idiotis"*, which in ancient times referred to a person who did *not* participate in the commons, i.e. the *polis* — in comparison with someone who did take part, called *"politis"*. Paradoxically, this word is used today with an entirely different meaning in several indo-european languages.

With the advent of Web technologies, privacy has become a topic of controversy. For instance, the popularity of social network sites has blurred the boundaries between public and private. People share personal content online, yet their audiences are not any more small and distinct, instead a number of different parties can access a single person's information. Instead of a complete separation between the public and the private, in the online world the relationship between these two spheres is less clear.

During the last decades, the public-private duality has been studied from different fields, such as sociology, psychology and law. This chapter presents different privacy-related theories before and after the Web. We present a review of the main theoretical approaches to privacy in social interactions before the Web and then introduce the concept

of online privacy. Finally, we review several theoretical approaches to privacy decision-making on the Web.

2.1 Us and them

The definition of privacy has always been a subject of debate; privacy is a rather elusive concept since it has neither static nor objective nature (Margulis, 2011). What one person regards as private, another person may regard as public and with time both may change their minds on the matter. The most common but also narrow definition of privacy is "the right to be let alone" employed by Warren and Brandeis at the end of the 19th century, who referred to press photography as a privacy issue (Warren and Brandeis, 1890). Today privacy violations coming from the practices of the press are often far more intrusive (e.g. the Leveson inquiry[1]).

Alan Westin, a pioneering privacy scholar, developed one of the most popular privacy theories. He regarded privacy as a state, "a voluntary and temporary withdrawal of a person from the general society" with four sub-states; solitude, intimacy, anonymity and reserve. His theory also suggested that "privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (Westin, 1967). Westin conducted over 30 surveys in 40 years (between 1978 to 2004) that investigated consumer attitudes to privacy. He classified consumers into three groups: the privacy fundamentalists (i.e. privacy concerned who choose privacy over benefits), the pragmatic (i.e. people who weigh the benefits against the costs of their privacy decisions), and the unconcerned (i.e. trustful and not concerned about privacy). Based on the outcomes of his surveys, approximately 25% of Americans belong to the first group, 57% to the second group and 18% belong to the unconcerned (Kumaraguru and Cranor, 2005).

A well-known approach towards privacy was taken a number of decades ago by Irwin Altman, a social psychologist, who viewed privacy as a *social process* and identified a set of basic characteristics (Altman, 1975).

- It can be distinguished between desired and achieved privacy. This highlights the fact that people do not often achieve their desired levels of privacy.
- Privacy is an interpersonal boundary-control process, and
- an input and output process. These two characteristics point out that it is control-based and regulated by human interactions.
- Privacy is an optimising process, i.e. there is potential for improving one's privacy.

[1] www.levesoninquiry.org.uk

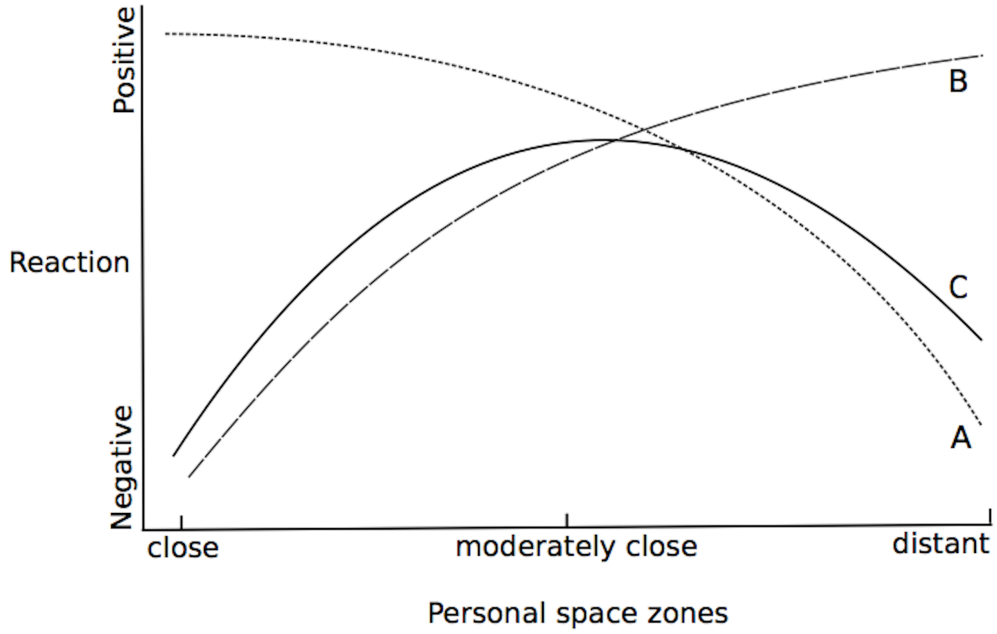
- It is a dialectic process. At a certain time the individual may want to interact with others, but other times they may want to be let alone.
- One of the most significant characteristics of privacy is that it is a *dynamic* process, in other words it changes over time.
- The last characteristic is that privacy involves a variety of social units, from the individual (self versus other people) to several social groups (one group against another) and so on.

Altman's approach to privacy is considered to be one of the best-articulated theories and has been extensively used in privacy research ever since. It differentiates itself from Westin's theory, as it focuses on the social interaction aspect of privacy (Margulis, 2011) and views privacy as a dynamic process rather than a state.

Sociologist Zygmund Bauman also focused on the interactions between the individual and others — *us* and *them* — and highlighted the potential privacy issues. To the eyes of each individual, people belong to one of three groups. The first group refers to people the individual interacts with on a frequent basis (friends and family), the second includes people that the individual meets on occasion (doctor, professor and so on), whereas the third group refers to all the people that the individual knows, but never meets. Any attempt from members of the second group to come closer to the individual and overcome the boundaries of their functional relationship can be regarded by the individual as a privacy breach (Bauman, 1990). Similarly to Bauman, Altman divides people to three groups: friends, strangers that the individual is expected to interact with, and strangers with no expected interaction. Altman uses a diagram to illustrate the expected interactions individuals may have with members of the three groups (Altman, 1975). Figure 2.1 represents Altman's diagram; curve A represents the reactions of the individual as someone from the first group approaches them (a friend), curve B the relationships with strangers the individual interacts with, and curve C the relationships with complete strangers (no expected interaction). As the diagram shows, the closer people from the last two groups attempt to come to someone's personal space, the more negative is the reaction of the individual.

In this context, Goffman stresses the importance of body language as a means of conveying information in people's interactions with one another (Gofman, 1963). In the case of privacy protection when the individual feels that someone enters his personal space he may express discomfort verbally but also through body language. The question that is now raised refers to what happens in a similar scenario on the Web. When the online dimension is put into play, the context collapses, as there are no physical interactions for the users to signal the discomfort provoked by a privacy breach. This raises the question of how people express discomfort in the online world taking into account that the traditional channels are absent. The principal way of expressing discomfort is by controlling

FIGURE 2.1: Altman's Graph on hypothesised relationships between personal space and reactions.



the access to their online personal space; this happens mainly through the management of their privacy settings, a topic that we will discuss later on in the following sections.

2.1.1 Concepts associated with privacy

Privacy is a term commonly employed by people in a variety of situations; however often people misuse the term in cases where they actually mean closely related concepts to privacy such as security, anonymity, or control. These concepts are associated with privacy, yet they are not synonymous to privacy. For instance, security is often confused with privacy. Camp discusses thoroughly the relationship between the two notions; privacy requires security, since without security the individual cannot control access to information about himself (Camp, 1999). Anonymity is also a notion confused with privacy. Although related they do not have the same meaning. Westin distinguished anonymity as a state of privacy - along with reserve, solitude and intimacy (Westin, 1967).

In fact, not only is there an epistemological difference between privacy, anonymity, and security, but also in practice they are opposing concepts. There is an on-going debate between privacy, anonymity and security that focuses on whether more security justifies less privacy and anonymity. The work of authorities would have been much simpler if there was no anonymity online, because cybercriminals would be easier tracked. On the

other hand, anonymity is a powerful asset for free speech. In a world without anonymity there would be many cases where due to ambiguities people could be exposed — Lessig discusses this topic thoroughly in his well-known book *Code v2.0* (Lessig, 2006). The SurPRISE project[2] re-examines the relationship between security and privacy, which is usually labelled as a trade-off. In cases where security measures and technologies involve data collection about people, questions are raised regarding privacy violations.

Privacy is often perceived as control (Margulis, 2011). For example, both privacy theories of Altman (1975) and Westin (1967) lie on the control of the individual over their information. Most scholars though argue that control is one of the factors that shape privacy (e.g. Smith et al., 2011; Margulis, 2011; Levin and Snchez Abril, 2009) and it should not be confused with privacy.

As already noted above, the relationship between privacy and all these associated concepts is often under debate. This is another reason why there is disagreement on accepting a single definition of privacy, and in fact it constitutes an obstacle in defining privacy (Margulis, 2003). In addition to this, the fact that privacy is studied from a number of different perspectives (law, psychology, economics and so on) that provide different insights make the acceptance of a single definition of privacy even harder (Pavlou, 2011).

2.2 Why is online privacy a subject of interest

Since the inception of the World Wide Web a number of privacy-related issues have arisen and as Web technologies continue to evolve these issues continue to grow. During the past decade the emergence of online social network sites allowed people to share a variety of information about themselves. In addition to this, the advent of the smartphone era has increased this sharing trend by allowing people to be almost permanently connected via these devices. Apart from the exciting opportunities that arose with the exchange of people's data in real time, concerns over people's privacy and safety came to the surface.

The previous section focused on privacy as a general concept. However, online privacy raises a number of new issues, as there is a lack of physical space (for example there are no bodily actions) and interactions are not face-to-face any more; instead they are mediated through the Web. Privacy decisions do not only have a short-term effect but a long-term one; information published on the Web are stored indefinitely (Palen and Dourish, 2003). Audiences can also be significantly large and *invisible*, in the sense that a number of different parties can access a single person's information without the necessity of a direct interaction with this person (boyd, 2008). Yet, the Web has become a part of our every-day life; we use the Web to communicate and interact with other people, to do our shopping, to access our bank accounts, to watch films, and many other

[2] surprise-project.eu

activities. In that sense, the boundaries between life online and life offline have become rather blurred (Turkle, 1995); we could even claim that there is no distinction between them any more.

Building upon Altman's theory (which was discussed in the previous section), Palen and Dourish (2003) developed a theory with a particular focus on online privacy. By reflecting on privacy's multidimensional and dynamic nature they considered three new privacy dimensions: the disclosure boundary (i.e. deciding what to disclose in different contexts), the identity boundary (i.e. managing self-presentation with different audiences), and the temporal boundary (i.e. current actions and their effect on future situations). It is important to note that these boundaries are dynamic and they change as the context changes.

This brings the question of what are the reasonable expectations of privacy in the online world. McArthur (2001) addressed this by highlighting that the social norms and the context surrounding a piece of information affect the extent to which the privacy expectations are reasonable. In the pre-Web era these expectations were more easily identifiable; it is reasonable to expect privacy at home but not in public spaces. McArthur argues that the World Wide Web is a transparent public environment; therefore expecting privacy in that domain is unreasonable. Similarly, Waldo et al. (2010) suggest that at a time when online surveillance is possible — and following the Snowden revelations it actually happens at a global scale — nobody should have a reasonable expectation of privacy. And as the famous quote from Scott McNealy, co-founder of Sun Microsystems, says *"You have zero privacy anyway. Get over it"*. Surprisingly, this quote dates back to 1999, yet several leading figures of tech giants have also made similar statements (e.g. Mark Zuckerberg and Eric Schmidt).

Still, most people wish to have control over their information (Coles-kemp et al., 2010) and the current legal framework looks out to support them (e.g. it supports people's expectations not to be under surveillance). In tandem with this, the Web has evolved into an entity where people publish a variety of information about themselves (e.g. details about their personal life in social network sites, their financial details, etc.). Boyd calls online social spaces *networked publics*, meaning that they are constructed through networked technologies and at the same time they are the communities emerging from the intersection of people, technology, and practice (boyd, 2008). O'Hara and Shadbolt (2008) call spaces, like the Web, that are neither entirely public nor private *privatised spaces*. Several studies have provided evidence that people do expect a certain amount of privacy even in these space (e.g boyd, 2008; Burkell et al., 2014), even though they do recognise that these spaces are not private. Cheung (2009) argues against the reasonable expectation of privacy and supports that it is rather insufficient to use as the standard in privacy protection given the nature of networked publics. Researchers also express concerns over people's excessive online social contacts and call this phenomenon *digital*

crowding (Joinson et al., 2011). As we will explore later on, *privacy in public* is rather important for individuals and there is significant research going on in this area.

This variety of information that people share online can potentially be stigmatising for them (Nosko et al., 2010). The findings from project KnowPrivacy from the University of California revealed that most of the top 50 websites collect personal data for customised advertising, and many of the well-known technology companies potentially share that data with hundreds of their affiliated companies (Gomez et al., 2009). This thesis raises the issue of the lack of awareness that people have regarding the actual affordances of their data. In addition to this, it explores the relationship between people's privacy attitudes and their actual behaviour. The following section is devoted in reviewing these topics.

2.3 Privacy decision-making

The mechanisms that people employ when making online sharing decisions are the main focus of this research. In the following section we present the theoretical background employed to approach these mechanisms.

2.3.1 The privacy paradox

The privacy paradox refers to the inconsistency that is often observed between people's sharing intentions and their actual disclosure practices (Norberg et al., 2007). The significance of the paradox lies on the fact that it raises a number of questions concerning people's online privacy behaviour, such as what are the contributing factors that make people decide to share information about themselves online and to what extent do people make these decisions rationally. The authors found that behavioural intention to disclose is influenced by risk but actual disclosure is not. This implies that perceived risk influences behavioural intentions but it is not strong enough to influence the actual disclosure behaviour.

Several studies have investigated people's paradoxical behaviour online. An experiment with a virtual shopping bot (conducted before social network sites became widely popular) found that although participants stated that privacy was important to them, they tended to disclose personal information to the bot (Spiekermann et al., 2001). A similar experiment confirmed the existence of the paradox and categorised its participants into four distinct groups; the privacy fundamentalists, the marginally concerned, the identity concerned (i.e. concerned about data such as their name and email), and the profile averse (i.e. concerned about revealing information such as their interests and hobbies) (Berendt et al., 2005). A more recent study that employed a mobile application, verified in practice that indeed there is a weak relationship between people's intentions

and actual disclosure behaviours (Keith et al., 2013). Through an online experiment where participants were initially asked privacy-related questions and then participated in an e-commerce scenario, Jensen et al. (2005) showed that apart from the fact that people have paradoxical behaviours, they do not have accurate perceptions of their own knowledge and understanding of online privacy.

At this point it should be noted that conducting research methods to study the privacy paradox is a challenging mission. Its study requires the investigation of people's privacy-related attitudes but also their privacy disclosure behaviour. According to Norberg et al. (2007), there are a number of challenges that are of particular relevance:

- Privacy perceptions vary widely among different people and are highly contextual.
- Different researchers use different research methods to evaluate privacy phenomena.
- So far, researchers have focused on privacy attitudes, intentions, and concerns, but not on the actual privacy disclosure mechanisms that individuals employ.

A number of studies have verified the existence of the paradox (e.g. Buckel and Thiesse, 2013; Keith et al., 2013). The above-mentioned issues, however, constitute significant limitations for conducting research in this topic. In fact, it is rather challenging to investigate people's privacy behaviour in the context of an experiment, simply because participants are aware that it is an experiment — even if they have been purposefully misguided by the researchers. For example, the fact that many experiments take place in a university may make the participants more trustful (Berendt et al., 2005). Taking all the above-mentioned issues into account, it appears that there is a number of challenges needed to be addressed in order to study the underlying reasons and mechanisms of the privacy paradox in online environments.

Some studies question the existence of a dichotomy between attitudes and behaviour for two main reasons. First, a few studies showed that there are cases where there is no discrepancy between people's attitudes and behaviours. For example, a pen-and-paper experiment that explored the monetary value of privacy found that participants with strong privacy concerns were willing to pay for their privacy, but they were also willing to accept more money for it (Grossklags et al., 2007). Studies exploring privacy in social network sites found that perceived privacy concerns do discourage people from online sharing, however these concerns can be mitigated when people feel in control of their privacy (Krasnova et al., 2010; Stutzman et al., 2011; Vitak, 2012).

Second, some scholars criticise the concept of the paradox as inaccurate. Preibusch (2013) suspects that the paradox is "an inaccurate interpretation of observable phenomena". Shklovski et al. (2014) used two separate studies — a series of interviews

exploring people's response to tracking and data leakage, and a survey focused on people's privacy attitudes towards data collection by applications — and found evidence for the privacy paradox. However, they questioned whether the privacy paradox may obscure the relationship between privacy disclosure and digital technologies. With that in mind they used the notion of *learned helplessness* to explain their findings. This notion is often employed to explain the process by which people come to accept certain situations by considering them as irreversible regardless of their own attempts to reverse them (Abramson et al., 1978). In this case, regardless of people's attitudes towards data collection and leakage, they still continue to download applications, because they accept that the situation will not change regardless of their own actions. Learned helplessness is an interesting notion, however it can easily be argued that people often use the argument "this is how things are" as an excuse to compensate for actually willing to trade their data in exchange for innovative services.

The privacy paradox serves a specific purpose as a concept; it raises the issue of the complex nature of privacy decision-making in our digital world. People do not have complete privacy simply by entering a private space online (e.g. their personal account in a web application), there is always an abstract feeling that there is an *invisible audience* that has access to one's private space. In fact, as we described earlier, this space is not really private; the Web is a *privatised* space. Potential entities that access one's space can be the system itself, a third party application or any other person that may gain access to that space. Due to this abstract nature of online space, attitudes towards online privacy vary; according to Westin (1967) people range from privacy fundamentalists to unconcerned. The privacy paradox sheds light upon the fact that people's sharing behaviours are often in conflict with their privacy attitudes. Certainly, the nature of their online behaviours is rather complex and the interplay between these behaviours and the way applications use their information is not trivial either, therefore they require further study. In tandem with this, as described earlier, some studies have shown that there are cases where privacy concerns may discourage disclosures. This also stresses the complexity of privacy behaviours. This thesis aims to explore the complex nature of privacy decisions and the reasoning behind them. In that sense, the paradox is an important concept because it acts as a starting point for the research presented here, as it raises all these issues, it highlights their magnitude and calls for further investigation.

2.3.2 The privacy trade-off

The second important concept regarding privacy decision-making is the so-called privacy trade-off, also known as the privacy calculus. The trade-off shows that privacy decisions are based on a cost-benefit examination (Acquisti, 2009); to what extent are people willing to release their data online in exchange for a particular service.

The privacy trade-off along with the underlying reasons behind the privacy paradox can be related to the ways people make consumer decisions, therefore studies on this topic have provided input into privacy research. Consumers make a cost-benefit examination to assess the outcomes of the release of their information (e.g. Hann et al., 2002; Hui et al., 2006, 2007; Dinev and Hart, 2006; Li et al., 2011a; Kehr et al., 2013). Bearing in mind that the tangible and intangible consequences of privacy are not easy to estimate, a number of factors may influence privacy decisions (Acquisti, 2009). According to Bettman et al. (1998), people are influenced by a set of preferences, which are often constructed on the spot, during the decision process — instead of predefined preferences. Predefined preferences usually come with experience with the decision environment (on the Web that could be a routine activity on a specific application), although situational factors may still influence decisions.

The privacy trade-off can be viewed as a form of social contract that governs privacy decision-making. The concept of social contract has long been employed to explain the expected behaviour during a transaction and its outcomes. The terms of the agreement are based on the assumptions that the involving parties acknowledge the existence of bounded rationality and recognise the need for a moral fabric (Dunfee et al., 1999). Fairness is fundamental in social contracts, as it empowers individuals with control and assures that all parties will adhere to the terms of the agreement (Culnan and Bies, 2003). In an online experiment investigating the willingness of participants to disclose information about themselves it was found that fairness (i.e. the collected data will be used for the intended purpose and nothing else) has a strong positive impact on people's disclosure decisions (Malheiros et al., 2013).

Acquisti and Grossklags stress a number of issues that help explain people's privacy decision-making. *Incomplete information* is a first issue, as people are expected to make privacy decisions with limited and asymmetric information — meaning lack of information about the possible outcomes after releasing one's data (Acquisti and Grossklags, 2005, 2007). For example, this lack of information can be on the access rights of third parties to a person's information but also on the individual's privacy protection levels from third party access (Acquisti and Grossklags, 2007). *Bounded rationality* (Bettman et al., 1998) presents an additional challenge to this (Acquisti and Grossklags, 2005, 2007), since people tend to make simplified decisions by choosing the path that requires the minimum cognitive effort but also minimum negative emotion instead of using mechanisms based on rationality. People also tend to focus more on the short-term effects of their privacy decisions rather than the long-term ones. *Hyperbolic discounting* refers to people's tendency to focus on immediate gratification rather than the potential long-term risks (O'Donogue and Rabin, 2000; Acquisti, 2004; Acquisti and Grossklags, 2004). For instance, a recent study by Microsoft on online reputation showed that only less than half of Americans think on the long-term consequences that their online activities may have on their reputation (Brackenbury and Wong, 2013). Hyperbolic discounting creates

a conflict between today's preferences and future preferences (Laibson, 1997). *Immediate gratification* is obviously a concept directly related to discounting, as it points out that people look for the immediate benefits of their trade-off. In that sense, people under-indulge in cases where the costs are immediate and the benefits are in the long-term (e.g. procrastination), whereas people over-indulge in the opposite cases because the benefits are immediate (O'Donogue and Rabin, 2000). Another privacy-related issue is the merger of different audiences into one single group, known as *context collapse*. It is a common characteristic among many web applications such as Facebook (boyd, 2008; Marwick and boyd, 2010; Vitak, 2012).

Several other psychological biases may play a role in people's decision making. Acquisti and Grossklags (2007) suggested a number of simplistic and paradoxical biases, such as overconfidence, the valence effect, rational ignorance, status quo bias, reciprocity and fairness, and inequity aversion. The *valence effect* refers to the belief that privacy invasions can only happen to other people but not to the individual himself, whereas *rational ignorance* appears in cases where deep understanding of a situation to form a rational decision requires much more effort than simply accepting its benefits. *Status quo bias* stresses the fact that people do not like changes and they prefer things to remain the same (e.g. not changing their privacy settings). *Reciprocity* and *fairness* refer to people's aspiration to act fairly in their trade-offs. Closely related to this concept is *inequity aversion*, which deals with people's discontent when they feel that others are unfairly getting rewards they do not deserve.

Other factors that may also influence privacy decisions are economic benefits (i.e. information in exchange for monetary rewards) (e.g. Hann et al., 2002; Hui et al., 2007) and incentives towards price discrimination (e.g. vouchers), given the condition that the information required to disclose is relevant to the purpose of the trade-off (Odlyzko, 2003; Li et al., 2010). Several studies that conduct experiments using monetary benefits show varying outcomes on privacy decisions. Although people tend to be willing to accept to give away their privacy in exchange for benefits, studies have shown that they are not as willing to pay in order to protect it (Grossklags et al., 2007; Acquisti et al., 2009). However, people's willingness to pay for privacy may change when there are clear indications of privacy protection. A study based on an online shopping experiment showed that people are willing to pay a premium to shop from websites that are more privacy protective (Tsai et al., 2010). Several studies have shown that when people feel comfortable and in control they are likely to share their information online. A study that used location-based coupons showed that TRUSTe[3] seals and legal statements on people's had a positive effect on sharing decisions (Xu et al., 2009). However, an earlier study had slightly different results; privacy statements made more participants to disclose their personal information but TRUSTe privacy seals did not (Hui et al., 2007). An experiment that provided people with feedback — regarding past requests from other

[3]Company that provides privacy seals to websites.

users to see their location — revealed that through feedback provision people felt more comfortable, and in turn more willing to share their data (Tsai et al., 2009). Brandimarte et al. (2012) ran a study consisting of three separate experiments that showed that when people have the perception of being in control over the access to their data, they tend to reveal more sensitive information, even though in reality they become more exposed. Wiese et al. (2011) ran an online survey where participants were asked to share information with others in several scenarios and found that when someone feels close to someone else they are more willing to share information with them. On the other hand, a study that used disclosure justification messages (i.e. justifying to the participants why they ask for their data) in an online experiment with a recommender system hoping to observe more disclosure, revealed instead that the justifications turned the participants to be less satisfied with their interaction with the system and did not increase disclosure (Knijnenburg and Kobsa, 2013a). Finally, a recent study with an experiment where participants bought a DVD from two separate online shops found that they were not willing to pay for privacy in cases where there was €1 discount at the shop that asked for personal information. In addition to this, they found that even in the case where there was no discount at the shop that requested personal information only half of the participants used the shop that did not (Beresford et al., 2012).

All these studies highlight a significant issue, which is that people do not make privacy decisions based on rational and consistent mechanisms (Acquisti et al., 2009). People trade their information based on a set of contextual and heuristically defined preferences, rather than a rational evaluation of the consequences of their decision.

2.3.3 Structuration theory and the privacy trade-off

The privacy trade-off as a process of decision-making may also be linked to a theory, developed by Anthony Giddens, known as structuration theory. Norberg's study showed that behavioural intention is governed by risk, whereas actual disclosure is not (Norberg et al., 2007), as described in a previous section, which can potentially be understood through the application of this theory.

Structuration is a social theory that frames behaviour as a balance between structure and agency, where the *structure* refers to the *rules* and *resources* that shape people's behaviour, whereas the *agency* refers to people's ability to act based on their free choices (Giddens, 1984). Giddens uses language as an analogy to *rules* pointing out that people react strongly when others disrespect the rules of a language. Similarly, people are assumed to meet their social expectations. *Resources* are the frames of reference where rules are carried out. They can be allocative, establishing control over things (e.g. someone's land) or authoritative, establishing command over others (e.g. someone's social status). Structure is seen by Giddens as a source of constraint on people's free

choices. In that way the number of choices that are available to people is constrained through structures.

The relationship between agents and structures has been the subject of a long dispute among sociologists. Bourdieu (1977) constructed a theoretical model of social practice in which he introduced the concepts of *habitus* and *field*. Habitus expresses the way in which agents develop attitudes and dispositions, but also the ways in which they engage in practices (Webb et al., 2002). In that sense, habitus is embodied in agents and it is an unconscious formation (Adams, 2006). A field is "a structured system of social positions — occupied by either individuals or institutions — the nature of which defines the situation for their occupants [...] a field is structured internally in terms of power relations" (Jenkins, 2002). As agents move across different fields, they incorporate into their habitus the structures of those fields. Bourdieu's theory offers a view in which agents do not really intervene in the way that the world works (Jenkins, 2002).

Giddens offered a new approach to this subject. At the core of the theory of structuration lies the notion of *duality of structure*; structures are the medium for decision making but at the same time they are the outcome of the agents' decision-making. In Giddens' own words "social structures are both constituted by human agency, and yet at the same time are the very medium of this constitution". In that sense structure has a dynamic nature that is reproduced through practice.

Structuration theory suggests three ways that structures become embedded in social interactions:

- Communication of meaning. People make use of *interpretive schemes* (e.g. a person who wears a white coat is a doctor).
- Use of power. People make use of resources involving *structures of domination* (e.g. a person who wears a badge is a police officer and has the authority to exercise certain power).
- Application of sanctions. People make use of societal norms involving *structures of legitimation* (e.g. people are expected to be formally dressed at a wedding).

According to Giddens, humans are reflexive individuals in the sense that we have different levels of consciousness that affect the way we behave in different contexts. In contrast to Bourdieu, reflexivity is an important feature of structuration theory. Giddens distinguishes two important levels of consciousness. The first level of consciousness is called *discursive consciousness*; at this level people are able to reflect and justify their actions and knowledge. The second is called *practical consciousness*; it includes all the cases where people know how certain things are in practice (knowledge of rules) but cannot describe them in a discursive way. For example, we take for granted that a dog is a dog and a cat is a cat. Practical consciousness enables people to go on with their daily

routines. Through the repetition of people's routines these structures are continuously recreated (the duality of structure). Therefore, the concept of practical consciousness is principal in structuration theory. At this point a question regarding privacy decision-making is raised; to what extent are privacy decisions part of discursive consciousness and to what extent are they part of practical consciousness?

This theory has the advantage of being a rather abstract theory; as a result it can be integrated into different contexts — although Giddens himself intended it to remain theoretical. Stones (2005) develops an approach called *strong structuration theory* that overcomes the abstract nature of the theory and places it *in situ*, and calls for empirical research to be conducted using structuration. "It (structuration) can focus on any set of surface appearances and make our understanding of them richer and more meaningful by elaborating upon the structures and agents involved and placing them in relevant networks of social and historical relations". In this research we use structuration in attempt to understand the privacy trade-off on the Web.

Giddens' theory has already been used and expanded to understand the relationship between technology and people. Jones and Karsten (2008) provide an in-depth review of papers in the field of Information Systems that employ structuration. Orlikowski (1992) introduced the "duality of technology", where technology obtains structural properties: it is the product of people and it is the people who apply a meaning to it, however when it is used in practice it becomes institutionalised. Adaptive Structuration Theory was also based on the work of Giddens and focused on the relationship between "information technologies, social structures and human interaction" in order to challenge what was perceived as a technocratic view of technology usage (Desanctis and Scott, 1994).

Although structuration has been applied to different technology domains, to the best of the author's knowledge it has not been used in the sphere of privacy on the Web.

With regards to online privacy, structuration theory would suggest that people are constrained by structures (such as trust or social expectations) when making privacy decisions and therefore do not act entirely as free agents. Here, we will go through a number of different examples where structuration could be applied. For example, the constraint could potentially account for the differences between their stated privacy attitudes, and their actual privacy behaviour. An example of such a structure could be the privacy settings that a specific online application applies as standard and every user is expected to follow (establishing a strong norm). At the same time though the privacy settings can be changed, depending on the feedback they received from people's use of them. Structuration theory could potentially explain the results of John et al. (2011), who found that a professional looking website raises more privacy concerns than an unprofessional site. The change of context had significant impact on the sharing decisions of their participants. A web-based survey on Flickr users showed that community-specific privacy concerns do affect people to choose more restrictive privacy settings, whereas

trust in other members and the community's information sharing norms decrease privacy concerns (Nov and Wattal, 2009). In both studies, the structures that govern the respective applications had an immediate effect on people's privacy concerns. In the first study through the *interpretive scheme* participants acknowledged whether these were professional looking websites or not, whereas in the second study the structures of *legitimation* showed that the sharing norms that govern Flickr expect users to perform certain actions (upload photos in this case).

This research aims to investigate whether there is evidence that indicates that the privacy trade-off can be related to the theory of structuration. In other words, this thesis will explore whether people are influenced by a set of structures when they make privacy decisions that causes them to deviate from their previously stated beliefs and whether those decisions could potential reinforce or create new influential structures.

2.3.4 Cognitive dissonance and the privacy trade-off

In the late 1950s Leon Festinger, an American social psychologist, developed a theory known as cognitive dissonance. Cognitive dissonance is a psychological state of tension that occurs whenever an individual is faced with a situation that is in conflict with their belief system. According to this theory, people are naturally driven to find a balance between their beliefs and actions to avoid the dissonance. Festinger suggested three paths that help people to overcome dissonance. The first one is by changing their beliefs, the second by changing their actions, and the third path by changing their perception of their actions (Festinger, 1962).

Festinger and Carlsmith (1959) tested this theory with a classic experiment in which all participants were instructed to perform a tedious task. However, at the end of the task some of them were offered $20 to describe the task to the next participant, who was in reality an actor, as an exciting experience. The remaining participants were only offered $1 for the same task. Naturally, due to the money offered to them the participants of the first group were strongly motivated to describe the task in a favourable way, in fact lie, whereas the participants of the second group were not. Surprisingly though, in the process of promoting the task as something interesting, the participants of the second group actually changed their own opinions and claimed that it was indeed interesting; thus showing signs of cognitive dissonance.

A simple example to understand this theory is smoking. Although people are aware that smoking causes serious health problems, they still do smoke:

"The person who continues to smoke, knowing that it is bad for his health, may also feel (a) he enjoys smoking so much it is worth it; (b) the chances of his health suffering are not as serious as some would make out; (c) he can't always avoid every possible dangerous contingency and still live; and (d) perhaps even if he stopped smoking he

would put on weight which is equally bad for his health. So, continuing to smoke is, after all, consistent with his ideas about smoking" (Festinger, 1957).

In the almost six decades of its existence, a number of different experiments using Festinger's theory have taken place. For example Brehm's well-known free-choice paradigm (Brehm, 1956), the experiment on the effect of punishment on cognitive dissonance (Aronson and Carlsmith, 1963; Aronson, 1997), as well as more recent studies on the origins of cognitive dissonance (Egan et al., 2007), and studies taking a neuroscience perspective (Jarcho et al., 2011; Izuma et al., 2010). Over the years the theory has matured through the various criticisms and revisions but also through the different areas it has been applied to (Cooper, 2007; Metin and Metin Camgoz, 2011). For the purposes of this thesis, we will try to view privacy disclosure decisions through a cognitive dissonance lens.

Similarly to the previously mentioned example, cognitive dissonance can be connected to the privacy trade-off. Individuals who share their data knowing there are potential implications to their privacy may also feel that they enjoy sharing aspects of their lives with their online friends, or that there is only a small likelihood that their privacy would be seriously violated, or that even if they stopped sharing at any moment, their information is already online (due to data persistence on the Web). In that sense, data sharing is consistent with their ideas about privacy.

In the case of privacy decisions, cognitive dissonance may occur in two separate cases. The first case deals with the *privacy paradox*, which was described in the previous section. The paradox stresses the existence of a discrepancy between people's privacy attitudes and their actual disclosure behaviour. In a similar way, cognitive dissonance deals with conflicting attitudes and behaviours. In the case of the privacy paradox, it may occur when people attempt to justify their disclosure behaviour when in discordance with their privacy attitudes. Festinger stresses that dissonance is a usual phenomenon when people form an opinion or make a decision; it is inevitable when they decide to act in one way but their knowledge and beliefs point to another direction. In that way cognitive dissonance can be a useful tool in understanding people's paradoxical behaviours in privacy decision-making. As we saw in the section on the privacy paradox, Shklovski et al. (2014) used the notion of *learned helplessness* to justify their findings and criticise the paradox. However, cognitive dissonance can explain the reason behind people's paradoxical behaviour; in the case of the study by Shklovski et al. (2014), the excuse that people accept that the situation will not change regardless of their own actions could easily serve a means to reduce the dissonance caused by their actions.

The second case deals with situations where people might have regretted their privacy decisions. In this case people decide to share their data online and later on they experience tension between their sharing behaviour and their beliefs. Regretting the sharing of their data can be seen as a means to minimise the cognitive dissonance that was raised

by this action. However, their action cannot be easily taken back, since their data may be stored indefinitely. Several studies on online sharing behaviour in social network sites have found that signs of regret are common when study participants are asked to justify their past online disclosure decisions (Sleeper et al., 2013b,a; Patil et al., 2012; Wang et al., 2011).

2.4 Conclusion

Privacy is a concept that has long been under study. Yet, with the success of the World Wide Web it has come to the surface again as an urgent research topic. This chapter began with a review of privacy theories prior to the emergence of the Web. An extensive review of a number of theories related to privacy decision-making along with empirical studies that explore these followed. These theories offer the theoretical background for the research presented in this thesis. The next chapter explores how privacy decision-making is a relevant topic to the Social Web.

Chapter 3

Privacy and the Social Web

With the emergence of web applications that enabled user-generated content and social interactions, the Web became a place where people perform a number of activities (from online shopping and gaming to posting personal details in social network sites). The sets of relationships that connect people across the Web constitute, what we call today, the Social Web (Halpin and Tuffield, 2010). Important components of the Social Web are social network sites; defined by Ellison and boyd (2013) as "networked communication platforms in which participants *a*) have uniquely identifiable profiles that consist of user-supplied content, content provided by other users, and/or system-provided data, *b*) can publicly articulate connections that can be viewed and traversed by others, and *c*) can consume, produce, and/or interact with streams of user-generated content provided by their connections on the site".

Privacy on the Social Web has been acknowledged as an area of significant concern especially when it comes to user awareness of privacy risks and their ability to manage personal information (Loukides and Gkoulalas-Divanis, 2009). It requires the utmost attention of privacy scholars as it brings out a new dimension to online privacy, and that is identity. It refers to the social identity people use on the Web, but also the identity that is constructed by the presentation of oneself. The Web is no longer as it used to be at its early days — an entity with a small number of users and even fewer content providers. Instead, it has grown to become part of everyone's daily routines, a place where people present themselves through content sharing and interacting with others.

This chapter provides an overview of the current state of privacy research on the Social Web. It explores a number of different topics, such as human interactions online and their relationship to privacy, the demographics of privacy, but it also hopes unveil the relationship between trust and privacy online. The chapter continues with an overview of empirical, but also technical approaches to privacy management.

3.1 Interactions in the Social Web

In the Social Web every user leaves their footprint that is stored permanently in digital format. O'Hara (2010) calls this tendency of personal information sharing *Intimacy 2.0*. Intimacy is an essential part of people's lives as it fulfils the human desire for relationship establishment. Social network sites provide an ideal ground for this, as they enable people — especially young people — to maintain friendships but also to explore their identities (Clarke, 2009; Livingstone, 2008). For instance, through a series of interviews Livingstone (2008) observed differences in the way teenagers present themselves in social networks as they grow older (from a more elaborate, stylistic identity at a younger age to a more plain identity focused on connecting with others in their late teenage years).

3.1.1 Self-presentation

Self-presentation is not a new concept and definitely not solely related to the Web; it was thoroughly investigated by sociologist Erving Goffman, who studied human behaviour in social interactions and affirmed that people wish to control the way others perceive them during their face-to-face interactions (Goffman, 1959). Goffman approached this topic with an analogy, more specifically by analysing theatrical performances. Similarly to theatrical performances, in social interactions each person is a performer who projects a conception of himself. In different contexts a person may give different performances. In addition to this, the performer tries to control the impression he gives to others. However, as the author points out, during each performance when the performer *gives* an expression at the same time he also *gives off* another expression to his audiences. In other words, the performer projects two different expressions of himself; their first expression is intentional, whereas the second one unintentional. In that sense, a process does take place, which consists of the following steps: a person gives a performance to an audience, the audience interprets his performance, and finally the person adjusts his performance based on the feedback of the audience. This process is called *impression management*.

The Web has offered an outstanding opportunity for people to present themselves to others and receive feedback on their performances. Similarly to Goffman's theatrical performances, on the Social Web people try to present a certain image of themselves and they often underestimate the fact that they give off a different impression to their audiences, which makes them look self-important (Barash et al., 2010). As studies have shown, other people (i.e. online audiences) do indeed influence and challenge a person's self-presentation in the Social Web in a way similar to offline interactions (Litt et al., 2014). Even before the Social Web, scholars explored the formation of identities through digital technologies. For instance, in the 1990s Sherry Turkle explored how multi-user

domains (multi-player real-time virtual worlds, mainly text-based) helped people explore aspects of their selves (Turkle, 1995). However, as boyd (2014) points out, in social network sites people communicate with others that they often know from physical settings and present themselves in contexts that are closely related to unmediated social communities.

In face-to-face performances the "personal front" of the performer includes a number of characteristics such as clothing, gender, age, and so on. In addition to this, as already stated in the previous chapter, body language may also be used to convey information. In online interactions, however, there is a lack of physical presence. Instead interactions take place through the exchange of different types of digital content (messages, photos etc.). The norm "one body, one identity" does not apply to the Web. In complete contrast to that, an individual may have multiple online personas (Donath, 1999). However, with the advent of social network sites and Facebook in particular, people started developing profiles under their real names (although fake profiles can still be created). This feature of social networks empowers people to construct digital identities and can be particularly useful for teenagers who are at an age of identity establishment.

Another difference between Web and physical performances is that the first can happen with great ease — with a click of the mouse — and may be projected to multiple audiences simultaneously. For example, people can easily show that they are at a specific location — and possibly reveal their activity in real-time as well as the people they are with — that would increase the potential for social interaction by simply doing a check-in through their social network account (Wang and Stefanone, 2013).

However, the lack of a visible audience complicates online interactions. The issue with *invisible audiences* is that often people underestimate the size of their audience (boyd, 2008). For instance, a study at the early days of Facebook showed that misconceptions about one's visibility within the network are common (Acquisti and Gross, 2006). A more recent study from Stanford University and Facebook followed 220.000 users of this social network site over a monthly period. They found that users hugely underestimated the size of their audience by a third of its actual size. In fact, each post reached instantly 35% of their friends and 61% of them over a month (Bernstein et al., 2013). Yet, in some cases people may decide not to share content online precisely because of the potentially large audiences, using self-censorship as a means to control their self-presentation (Sleeper et al., 2013a).

Self-presentation is often related to narcissistic traits and social network sites provide a new window of opportunity for people to promote themselves (Carpenter, 2012). A study comparing users with non-users of Facebook revealed that indeed users tended to be more narcissistic than non-users (Ryan and Xenos, 2011). In that sense, social network sites are ideal spaces because people can have hundreds of connections (hence a large audience) and be in complete control of the content they post (e.g. photos of themselves)

(Mehdizadeh, 2010). Studies have also revealed that the ways through which narcissists portray themselves in these spaces are similar to the ways they portray themselves offline (Buffardi and Campbell, 2008). Their main characteristics are higher levels of social interaction and more frequent publication of self-promoting content (Mehdizadeh, 2010; Ong et al., 2011), methods that can also be employed offline.

3.1.2 Social capital

The feedback people receive through their online interactions is not only related to self-promotion but may in fact offer them a number of significant benefits. Such benefits can be emotional support, advice, new information and ideas and so on. For instance, Vitak (2012) studied self-presentation in social network sites through the form of self-disclosures (e.g. status updates) and stressed that it has an important role for relationship maintenance. All these benefits are part of what is called *social capital*. Bourdieu (1986) describes social capital as the total of the actual or potential resources people may benefit from through their social network. Ellison et al. (2011) note that it can be viewed as a form of capital embedded in social relationships. Since social capital deals with social relationships, it can be of benefit to more than one person (Schmid, 2000). In fact, it may offer benefits not only to individuals but also to social groups (Lin et al., 2001).

Social capital is a long-studied concept in sociology — key contributors include Bourdieu (1986) and Coleman (1988) — but it has also been explored in other research areas such as economics, politics, and management (e.g. Adler and Kwon, 2002; Sander, 2002; Putnam, 2000; Nahapiet and Ghoshal, 1998). As a consequence, numerous definitions of this concept exist (e.g. Adler and Kwon, 2002, have listed several key definitions from different scholars), yet they all agree that social capital deals with social relations and their potential benefits. Coleman (1988) highlights the beneficial nature of social capital by stating that it is productive, since it enables the achievement of certain ends which would not have been made possible without it. Nahapiet and Ghoshal (1998) distinguish between three dimensions of social capital: the structural (i.e. the structure of the social network), the relational (i.e. the social relations), and the cognitive dimension (i.e. the shared interpretive schemes of the people within the network).

Apart from the lack of a commonly acknowledged definition there is also a lot of controversy over the appropriateness of the use of the word “capital” in the term. Often researchers point out a clash between its usage in the concept of social capital and the classic meaning of the word “capital” in economics. According to Smith and Kulynych (2002) this word offers a too broad, pervasive, and honorific meaning to the concept. Similarly, an issue for economists is that it cannot be measured in monetary terms (Solow, 2001; Piazza-Georgi, 2002). A number of different scholars have debated over

this, some have even offered alternative terms to replace it, but this debate is still ongoing (e.g. Piazza-Georgi, 2002; Smith and Kulynych, 2002; Lin et al., 2001; Portes, 2000).

Several researchers have focused mostly on the positive aspects of social capital. However, it may also have undesirable effects. Portes (2000) points out some potential unwanted outcomes such as exclusion of outsiders (i.e. it has benefits only for people within the group), excess claims on group members (i.e. a free-rider problem emerges), restrictions on individual freedoms (i.e. members need to conform to the group), and downward levelling norms (i.e. not promoting social mobility in order to sustain group cohesion).

This thesis focuses on the role of social capital in the Social Web in particular. During the past few years a number of studies have focused on the existence of social capital in online environments and especially in social network sites. Indeed these spaces offer a perfect ground for building and maintaining one's social capital (Hampton et al., 2011; Steinfield et al., 2008; Chiu et al., 2006). Studies often make a distinction between two forms of social capital, *bonding* and *bridging* social capital (Vitak, 2012; Burke et al., 2011; Ellison et al., 2007; Williams, 2006). Putnam (2000) presented in detail these two distinctive kinds of social capital in his book "Bowling Alone: The Collapse and Revival of American Community". The first refers mostly to the emotional support offered by close relationships (strong ties); research has shown that we increasingly use social network sites to keep up with people close to us. A survey of the Pew Research Centre found that Internet users and especially social network users get more social support than non-users — in terms of emotional support, companionship, and tangible support (Hampton et al., 2011).

Bridging social capital refers to access to new information that is offered by weaker connections; as Granovetter (1983) suggested weak ties act as bridges between network segments. For example, a Facebook study showed that people engage themselves in information seeking strategies to learn more about acquaintances that they know little about (i.e. their weak ties) (Ellison et al., 2011). Burke et al. (2011) found that different types of activities on Facebook can have different effects on bridging social capital by showing that receiving messages from friends is associated with an increase in bridging social capital, whereas other activities do not have similar effects (the other activities they studied were passive reading of news, and simple status updates). Yet, the study also showed that passive reading of other people's news does offer support to people with poor communication skills. Other studies have indicated that social network sites may help people with low self-esteem (more than people with high self-esteem) to overcome their barriers and build large networks — which are sources of bridging social capital (Steinfield et al., 2008). However, sometimes this can happen by sharing exaggerated information about themselves (Zywica and Danowski, 2008).

There is a lack of research regarding the interplay between privacy attitudes, disclosure behaviours and social capital. Stutzman et al. (2012b) also highlighted this gap and studied this triadic relationship using an online survey targeted to social network users. Their findings suggest that privacy has an indirect relationship with social capital, as it influences directly disclosure behaviours, which in turn bring out social capital outcomes. This thesis will address the role of social capital in disclosure decisions and privacy management in Chapter 6.

3.2 The demographics of privacy

A significant amount of research has focused on questions that are raised on people's attitudes and behaviour online with regards to privacy. Social network sites, in particular, have provided new ground for research both as a topic of study but also as a medium for data collection.

Many studies have focused on different demographics (e.g. age, gender) and their impact on people's privacy attitudes and behaviour. For instance, a topic of great interest and debate is the use of online services and especially social network sites from young people. The following sections explore differences in privacy attitudes based on demographics.

3.2.1 Gender differences

Both men and women use the Web on a regular basis to a similar degree. According to Pew Research Centre, in 2011 80% men and 76% women in the USA were on the internet (Zickuhr and Smith, 2012), percentages that by 2013 escalated to 85% and 84% respectively (Zickuhr, 2013b).

With regards to online privacy, research has often provided evidence that women tend to have more privacy concerns than men (Wills and Zeljkovic, 2011; Hoy and Milne, 2010; Fogel and Nehmad, 2009; Youn and Hall, 2008), and in fact are more likely to be more private online. Several studies show that women who use social network sites are more likely to have their profiles set as private compared to men (Madden, 2012; Lewis et al., 2008). An early study on Facebook (Facebook was launched in 2004 the study took place in 2006) showed that women at the early years of this social network did not share specific information about themselves, such as their sexual orientation, address and phone number (Acquisti and Gross, 2006). A more recent study that crawled 479K Facebook profiles found that men share more publicly personal information in comparison with women, yet the difference between them was small (Farahbakhsh et al., 2013). Another study found that men and women of equal digital skills tend to equally share their content online (Hargittai and Walejko, 2008). Still, Quercia et al. (2012) found that women are more likely to share information about themselves less publicly

than men. Significantly, women are more likely to take certain measurements to protect their privacy (Thelwall, 2011). For example, they are more likely to read privacy policies before creating an account in a social network site, "untag" themselves from photos, and be careful about whom they friend (Hoy and Milne, 2010). However, both men and women are equally likely to search for their digital footprint (Madden and Smith, 2010).

3.2.2 Young people

A long debate has taken place regarding the attitudes towards online disclosures and privacy management from different age groups. For a long time it was a common claim that young people do not care about privacy. Prensky (2001) coined the term *digital natives* to refer to young people that grew up in the current digital era — also known as the net generation (Tapscott, 2009, 1998) among other terms — distinguishing them in that way from older generations, known as *digital immigrants*. Palfrey and Gasser (2008) differentiated these two groups between people who were born after 1980 and grew up with the Web and people who were already adults at the time. To the eyes of the digital immigrants online privacy is viewed as a complete oxymoron. Parents and educators are the ones who should offer the appropriate guidance to younger generations; however the problem of digital literacy in the older generations constitutes an obstacle to its realisation. As a consequence, the solution to the issue of guidance often lies on younger generations themselves (Palfrey and Gasser, 2008).

The use of the terms digital immigrants and digital natives has often been criticised by the research community. Several researchers disagree with their use and state that there are multifaceted factors that affect people's digital skills, hence it is fundamental to conduct empirical research aiming to understand people's online attitudes and behaviour (e.g. Buckingham, 2008; Hargittai, 2010; Jones and Czerniewicz, 2010; Jones et al., 2010; Margaryan et al., 2011; Koutropoulos, 2014).

Recent research has also found that the claim that young people do not care about privacy does not hold in practice (Blank et al., 2014; Marwick et al., 2010; Hoofnagle et al., 2010; Christofides et al., 2011). Sanchez Abril (2007) use the analogy of car drivers to illustrate the fact that the so-called digital natives do expect a certain level of privacy even though they are in a public online place. Reports on social media usage show that nowadays teenagers and young adults have more private settings on Facebook compared to the past, however they do share a lot of information about themselves with large networks of online friends (Christofides et al., 2011; Stutzman et al., 2012a; Madden et al., 2013; Tessem and Nyre, 2013). Surprisingly, most teenagers are not very concerned about third-party access to their data (Madden et al., 2013). A Facebook study on young students (aged between 18–25) found that they care to manage their social privacy (i.e. how others view them), but they have little concern about institutional privacy (i.e. what applications do with their data) (Young and Quan-Haase, 2013).

3.3 The role of trust in privacy decision-making

This section focuses on the role of trust in the Social Web and especially on its relationship with privacy concerns and disclosure behaviours. Because of this, we look at trust from the perspective of the users; trust emerges when users believe that a specific web application is competent, benevolent and honest in handling their personal information (McKnight et al., 2002).

Beldad et al. (2011) suggest several aspects that may influence people's trust in an application; the existence of a privacy statement (even if they do not actually read it), the use of security measures, and its reputation. A study on people's willingness to share information with applications (with a focus on Foursquare[1]) found that frequent users are more confident and more inclined to share information (Tessem and Nyre, 2013). Other studies have reached similar results; they all agree that the frequent usage of social network sites has positive effects on people's confidence with online sharing, as well as their trust in these applications (Fogel and Nehmad, 2009; Frye and Dornisch, 2010; Lin and Liu, 2012).

The relationship between these three constructs — privacy concerns, trust, and disclosure behaviours — is rather complex and delicate, as pointed out by several researchers (e.g. Joinson et al., 2010; Taddei and Contena, 2013). When Norberg et al. (2007) introduced the concept of the privacy paradox they also started their experiment with the hypothesis that behavioural intention to disclose is influenced by risk, whereas actual disclosure is based on trust heuristics (i.e. trust has a direct influence on privacy behaviour). However, their results did not offer sufficient evidence to support the hypothesis that trust does influence actual behaviour. Since then, there has been a substantial number of studies (mostly survey-based and some experiment-based) that has provided evidence that trust does indeed decrease privacy concerns (Joinson et al., 2010; Taddei and Contena, 2013) and influence in a positive way sharing intentions (Lin and Liu, 2012; Zimmer et al., 2010; Dwyer et al., 2007), and disclosure decisions (Mesch, 2012; Krasnova et al., 2010). However, studies have come out with contradicting results regarding the extent of that influence; whether trust influences directly or indirectly sharing intentions and behavioural disclosures (Taddei and Contena, 2013).

Krasnova et al. (2010) found that perceived control over one's information also plays a role in privacy decision-making, as the sense of control has a positive influence on trust, which in turn reduces perceived privacy risk. In conjunction with this, an earlier study found that risk awareness reduces trust and increases the demand for control (Olivero and Lunt, 2004). The relationship between trust and control was also confirmed by Taddei and Contena (2013), yet their study found that perceived privacy risks do not affect disclosure behaviours directly.

[1] foursquare.com

Taking all the above-mentioned into consideration, the role of trust in privacy decision-making appears to be complex, and therefore requires further analysis, as studies often come to different results regarding the importance of its role. However, it is commonly acknowledged that trust does indeed play a role in the privacy decision-making process. This means that systems need to provide an environment where a sense of trust does emerge.

3.4 Privacy management issues

A number of studies have focused on privacy management in social networks. They depict a mixed picture of users who are concerned about privacy but struggle to set appropriate preferences and whose behaviour does not necessarily follow that concern. For example, in a US-nationwide survey 58% of the participants stated that they have private profiles in social networks, and at the same time 50% of the participants expressed difficulties in managing their privacy settings (Madden, 2012). A survey focusing on Facebook that was conducted twice with the same group of young people (in 2009 and 2010), revealed that being a regular user often coincided with more regular changes in their privacy settings (boyd and Hargittai, 2010). Another study showed that people who have a personal experience of privacy invasions are more willing to change their privacy settings than others (Debatin et al., 2009). A seven-year long study on Facebook in the USA found that with time people share more online, yet they do make their content more private. More specifically, the study collected Facebook profile data (such as home town, birth date, contact information, and interests) and found that with time the amount of profile data displayed *publicly* on the network decreased. Apparently, as time passes Facebook requests more data from its users, but the users are less likely to show their content to strangers (Stutzman et al., 2012a).

Still, the study revealed another major issue; with the current state of things all third parties (Facebook, advertisers, third party applications) collect even more data about the users (Stutzman et al., 2012a). To make things worse, there is a lack of user awareness regarding third party access to their data as well as potential data manipulations (Krishnamurthy and Wills, 2008). For instance, the TRUSTe Internet of Things Privacy Index found that only 47% of people in the UK [2] and 59% of people in the US [3] know that smart devices can collect details about their personal activities. In addition to this, the latest Pew Internet Report found that most American teenagers do not have strong concerns regarding third-party access to their data (Madden et al., 2013), which indicates the existence of another paradox. As systems and other third parties continuously collect information about people, people themselves are either unaware of these

[2] http://www.truste.com/gb-internet-of-things-index-2014/
TRUSTe Internet of Things Privacy Index - GB Edition

[3] http://www.truste.com/us-internet-of-things-index-2014/
TRUSTe Internet of Things Privacy Index - US Edition

affordances of their data or they are not even concerned about this issue and continue exchanging their information with these systems.

Another question that needs to be addressed refers to whether people are indeed satisfied with the current privacy settings. The main reason behind this is that people's internal privacy preferences often do not match the ones offered to them (boyd and Hargittai, 2010; Madejski et al., 2012). In other words, the ways in which systems expect people to express their privacy preferences, which are mostly access-control based, do not match people's actual privacy decision-making. A Facebook study showed in practice the mismatch between participants' sharing intentions and potential privacy breaches by comparing their reported privacy preferences, their actual privacy settings and their posts (Madejski et al., 2012). Another Facebook survey focused on people's privacy settings and the audiences of photos they posted online also highlighted the issues people are having with managing their privacy settings. The study revealed that approximately half of the photos were posted with the default settings making them visible to everyone on the social network; even in photos where the settings were modified, only 37% of the time did the settings match the expectations of the participants and often the audience of the photos was bigger than expected (Liu et al., 2011). A study on Google+ showed that 85.7% of the participants (who are active Google+ users) occasionally share posts *publicly* (Kairam et al., 2012). This means that the issue of privacy management of people's day-to-day disclosure behaviour in different applications — e.g. Facebook posts, location check-ins, tweets with location — is rather complex, therefore it requires a more analytical approach. It also poses challenging questions for researchers and practitioners on how to develop privacy systems that can capture people's actual privacy preferences. In order to succeed in that it appears indispensable to study the underlying mechanisms of people's privacy decision-making.

3.5 Approaches to privacy management

A number of initiatives have approached the issue of online privacy from different areas of computer science (e.g. security engineering, machine learning, human-computer interaction). One way of classifying the various studies is through three research paradigms; privacy as control (e.g. privacy settings), privacy as confidentiality (e.g. anonymous communications), and privacy as practice (e.g. transparent systems) (Diaz and Guerses, 2012; Danezis and Guerses, 2010). Apart from the technical approaches, policy makers and authorities have also taken steps to preserve people's privacy through legal processes. Several of these approaches apply directly to the Social Web and we will look into them in the following subsections.

3.5.1 Privacy settings

In recent years a common approach to privacy management is *audience segregation* (van den Berg and Leenes, 2010). Several studies have developed tools aiming to assist users at grouping audiences and using separate privacy settings per group (Mazzia et al., 2012; Amershi et al., 2012; Egelman et al., 2011; Lipford et al., 2010, 2008; Reeder et al., 2008; Adu-oppong et al., 2010). For example, a study where participants were offered choices to deny-or-allow access to certain people to their online posts (e.g. a person from another group who potentially should not see the post) assisted them in dealing with fewer audience issues (Egelman et al., 2011). Some experiments also use justifications, which inform the user about the privacy preferences of other users (their online social circle), hoping to assist the user in making better informed privacy decisions (Besmer et al., 2010; Patil et al., 2011).

A more dynamic approach to the issue of privacy settings is offered by studies that employ recommender systems based on machine learning techniques. Such an example is the approach developed by Fang and LeFevre (2010), who used a machine learning algorithm that requires minimum user input, instead it is mostly based on community characteristics (e.g. studying the emergence of different communities on the user's network, looking at profile data of friends of the user). Similarly to this approach, Li et al. (2011c) developed SPAC, a tool that makes inferences about privacy preferences based on the user's profile and past privacy settings. Ghazinour et al. (2013) developed a recommender system, called YourPrivacyProtector, which uses collaborative filtering based on the similarity between the privacy settings of the user and those of other users. Li et al. (2011b) developed a dynamic trust-based system that uses community detection algorithms to identify trust relations and enables users to choose privacy preferences (audience selection) on-the-fly to their posts.

All these studies are useful, as they provide deeper insight into the technical nature of privacy management. Yet, many of them lack an understanding of the nature of privacy decision-making and the reasonable expectations of privacy; therefore a more holistic approach to privacy is necessary. In addition to this, the above-mentioned systems address the issue of privacy preferences, but do not focus on disclosure behaviours, which are highly dynamic and contextual, as we will discuss in the following chapter.

The problem of privacy becomes even greater when applications use automatic disclosure of information about the users (such as someone's location as part of their post). According to Vihavainen et al. (2014) users are faced with three issues because of this; lack of sensitivity to situational factors, insufficient control over the specifics of the disclosed content as well as complete lack of control over disclosure to service providers and third parties.

Another approach to privacy management is offered by *nudges*. Nudges are a method of designing systems that provide information to users so that they can make informed privacy choices without diminishing their freedom (Acquisti, 2009). Such an approach would make people aware of potential biases or cognitive overload that can influence their decisions (Balebako et al., 2011). Wang et al. (2013) introduced three nudges and tested them on Facebook: the picture nudge, the timer nudge and sentiment nudge. The picture nudge showed user profile pictures of people that would be in the audience of the post, the timer nudge called for a time delay until the post was actually submitted, whereas the sentiment nudge displayed a notice based on the overall sentiment of the post. The nudges appeared in real time when users made a privacy decision taking into account unexpected audiences, and user regrets that people often experience after they have posted something. Among the three nudges (tested with a small set of participants) the picture nudge was the most successful, as it managed to address rather successfully the issue of *invisible audiences* without causing annoyance to the users. It should be noted that at the time of writing this thesis, Facebook itself announced the release of a "privacy checker" feature that prompts users to review their privacy settings (e.g. check the audiences of their posts) [4].

One of the advantages of nudges is timing. It has been shown — in the context of a research experiment though — that timing of privacy-related information may impact privacy decisions (Egelman et al., 2009). An experiment that used privacy notices followed by misdirections (e.g. in the form of time delays) instead of allowing participants to make their privacy decisions immediately, found that these misdirections reduced the impact of the privacy notices in the decisions (Adjerid et al., 2013). Although nudges have obvious advantages to people's choices, they do not offer a personalised and context-based solution, but instead a rather generic solution (Knijnenburg, 2013).

3.5.2 Transparency and accountability

Transparency and accountability are often proposed as a means of addressing online privacy. Instead of focusing on methods of personal information concealment, it is suggested to make the information transparent to everyone (O'Hara and Shadbolt, 2010). Transparency provides a useful solution when coupled with information accountability, in other words the transparent use of information in order to determine whether it is handled appropriately as defined by a set of rules that hold the individual accountable of possible misuse (Weitzner et al., 2008). Transparent and accountable systems enable not only authorities and institutions but also individuals to be aware of the online information flow. In his book *The Transparent Society*, Brin (1998) argues that people wish to see what others are up to, yet they are not willing to hold others accountable

[4] http://www.telegraph.co.uk/technology/facebook/10849164/Facebook-in-new-privacy-push.html/ Facebook in new privacy push, The Telegraph

of their own activities. As a consequence, transparency can offer a fair means of accessing information, since access to information is universal. There is still one challenge to overcome, since systems usually enforce accountability on users but not on "power structures" (e.g. authorities, organisations). According to the author, a possible solution is the development of technologies that enforce accountability to power structures and prevent them from enforcing accountability to citizens. Solove (2004) criticised this approach arguing that to make this feasible, people need to have the same capacities as power structures.

Transparency principles have been proposed from a number of different stakeholders; technical experts (e.g. Weitzner et al., 2008), legal scholars (e.g. Solove, 2004), but also authorities — Reding (2011) in the European Commission, and the Privacy Bill of Rights by the Obama administration. Several countries have also made significant steps by putting forward transparency programmes (e.g. data.gov.uk and data.gov). These programmes promoted the release of public government data online to be accessible by anyone interested. Transparency is also one of the seven foundational principles of *Privacy by Design*, a well-known initiative that aims to ensure that institutions operate with respect to people's privacy (Cavoukian, 2010).

We focus on the prospects of using transparent methods in privacy decision-making on the Web. Transparency and accountability can, to a certain extent, address the issue of *limited information* and help users make better informed privacy decisions. These can be fruitful approaches, as they ask for people's *informed consent* by making them aware of others who may have access to data about them, potential data manipulations, and data about others that the individuals themselves may encounter (Pötzsch, 2009). Still, their privacy decisions may suffer from other biases such as hyperbolic discounting and bounded rationality. As we also described in the previous section, simple changes (such as delays) in the framing of privacy notices, can reduce significantly the impact of the privacy notices in people's privacy decisions and in some cases they may even make the participants disclose more information than intended (Adjerid et al., 2013). Apart from this, even when information about data manipulation practices is disclosed appropriately to the users, privacy notices do not offer a practical solution due to their level of detail, and subsequently end up having the opposite effect than the intended one. This is what Nissenbaum (2011) calls the *transparency paradox*. Studies have found that reading privacy policies is extremely time consuming and in a language that not all users can understand, therefore impractical (Mcdonald and Cranor, 2008; Jensen and Potts, 2004).

Another problem with transparency is that it is often used as an alibi for data collection. The problem stems from the fact that the vast amount of data about people that are stored persistently online can be manipulated in several ways (by the systems themselves or other third parties) beyond user knowledge. For instance, a major concern is the fact that through online data, a number of inferences can take place to reveal more data

about people. To illustrate this with an example, a recent study found that based on easily accessible data (such as Facebook Likes) a number of sensitive information can be automatically inferred such as ethnicity, sexual orientation, political views etc. (Kosinski et al., 2013). As Nissenbaum (2011) argues, for a given moment only a snapshot of the information flows can be grasped, since these flows are in a constant change (due to the emergence of new analytics and services etc.). In that sense, the affordances of people's data make the issue of transparency and accountability even more difficult.

Transparency is a fundamental principle, as it assists individuals into making informed privacy decisions as well as it allows them to be held accountable of the practices of institutions. However, is not self-sufficient; instead it should be combined with other technical and legal approaches (e.g. Privacy by Design). For example, Garg et al. (2013) point out how transparent privacy mechanisms for information sharing can take place by viewing privacy as a community good that needs to be preserved through community-established norms.

3.6 Conclusion

This chapter was devoted to a review of on-going privacy research on the Social Web. The Social Web has turned into a new disembodied digital social network that mirrors the traditional, physical, one. It provides a space for social interactions through which people can explore their identities and increase their social capital. Privacy on the Social Web is a topic of paramount importance for the research community and society in general, since people have the tendency to share personal aspects of their lives without a proper awareness of the affordances of their data. In that sense, people do not take into account what happens with their data once they have released it (e.g. third party tracking issues). This issue becomes even greater, as current privacy systems do not support the actual privacy-decision process. As a result, people are likely to make poorly informed sharing decisions, and in turn their privacy is often violated without their knowledge. To that purpose, we discussed two initiatives that can assist people in making better-informed privacy decisions. Transparency holds the users accountable of the affordances of their data, whereas more sophisticated privacy settings provide a better support to the users in managing their audiences.

Chapter 4

Privacy and the Social Web: A Framework for Analysing Location Data

Over the past few years a new trend of location sharing has emerged. Applications prompt users to share their real-time location in a variety of ways, such as posts or photos in social network sites tagged with one's location or in exchange for innovative services (e.g. map navigation). The success of smart enabled devices was fundamental in this, because they offered applications the opportunity to access people's location through their GPS coordinates. 74% of smartphone owners in the United States get directions or other information based on their real-time location (Zickuhr, 2013a). Along with the benefits of location sharing strong concerns about people's privacy arose.

This chapter focuses on location information on the Social Web and the emerging privacy issues. The main argument that we build is that location privacy should be considered as part of a greater contextual or situational privacy. The first section provides a discussion and a background literature on this topic, whereas the second section presents a survey of technical systems that use location data. The third and final section introduces a conceptual model regarding the current state of privacy by bringing together the outcomes of the survey along with background research.

4.1 Location privacy and the Social Web

Current technologies on the Social Web offer people the opportunity to share their location data in real time; this includes features of social network sites such as Facebook (geolocated posts, photos and so on), Twitter (tweets with one's location), but

also location-based applications, such as Foursquare. The popularity of these applications has grown significantly in the last decade. According to the latest reports from the Pew Research Centre, 40% of Americans access social network sites through their smartphones[1]. These devices play a fundamental role in the development of such applications, as they are able to track their owner's location traces through their GPS coordinates. 74% of smartphone owners in the United States share their location either in exchange for directions or information based on location (Zickuhr, 2013a). In addition to this, there is a notable growth in the number of social media users who manage their accounts to include location in their posts — 30% of users (Zickuhr, 2013a). Research has also found that more active social media users are also more likely to share their location data, among other data about them (Tessem and Nyre, 2013). As studies regarding online location sharing reveal, the underlying motivations for this trend are strongly related to self-presentation, the opportunities to connect with other people but also to gain access to certain services — e.g. collecting vouchers (Patil et al., 2012; Lindqvist et al., 2011). This is also confirmed by statistics that reveal that 56% of young people are willing to exchange their location for coupons and deals (Lebo, 2013). Wang and Stefanone (2013) argue that location sharing "involves the announcement of a simultaneous presence of locations, activities, as well as social actors, thus increasing the potential for social interaction and impression to be developed". This trend is expected to continue to increase in the forthcoming years.

Along with the many advantages of these technologies a number of serious concerns emerge. The exposure of the exact geographic location of people poses privacy concerns, as it reduces significantly their anonymity (Karpf, 2009). With the advent of mobile Web applications concerns with regards to a special type of privacy have arisen, known as location privacy. Location privacy on the Web deals with the issues that may appear when people's location data are released to Web applications. It has been defined as a particular type of information privacy that focuses on the need of individuals to decide when and how others may access their personal location information (Duckham and Kulik, 2006). From a technical standpoint, location privacy deals with the capability of a mobile node (i.e. mobile device or router) to conceal the relation between the location information of the device and its personal identifiable information from third parties (Liu, 2009). At this point it should be clarified that when we talk about location, we talk about a category of information that encompasses different levels of accuracy; from GPS coordinates (i.e. high accuracy) to a semantic name of a place, such as the name of a city (e.g. London) or a more "personal" name of a location (e.g. John's restaurant). Still, location data always refers to a specific area on the map no matter the level of accuracy or the semantic ambiguity of this data.

Location privacy has been studied in the field of Ubiquitous Computing and over the last years in Participatory Sensing (Burke et al., 2006), also known as Social Sensing

[1] http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

(Aggarwal and Abdelzaher, 2013), an area that focuses on data collection through the combination of social networks and sensor applications (e.g. GPS applications). The successful spread of social network sites in smartphone and tablet devices has turned location privacy to a pressing issue in the Social Web. Enthusiastic users of such applications have the tendency to over-share information, giving away a plethora of easily available information about themselves (Rose, 2011). Due to this over-sharing of information, people's location can now become traceable, often without them actively giving it away. For example, Mahmud et al. (2014) developed an algorithm that uses a set of statistical and heuristics classifiers to infer Twitter users' home location (at different granularities) based on the content of their tweets (e.g. looking for names of locations) and tweeting behaviour (number of tweets per time may indicate time zone). In addition to this, research has shown that people can potentially be tracked through their smartphones and tablets at all times; even in cases where the location settings in their devices are off (Dey et al., 2014). Their sensors offer a means of tracking, as they leave fingerprints, which act similarly to browser cookies, due to some flaws in the hardware manufacturing process. Significantly, the Electronic Frontier Foundation developed an application that tested how unique is a user's web browser based on a set of characteristics (browsers plug-ins, time zone, screen resolution etc.), and revealed that in fact browsers leave their own fingerprints that can potentially be unique (Eckersley, 2010). These studies show in practice that from the moment people use a digital device they constantly leave their fingerprints through various ways, offering ground for potential tracking mechanisms.

4.1.1 Privacy attitudes towards location sharing

A number of studies have focused on privacy attitudes as well as on the privacy settings of web applications. Benisch et al. (2011) studied people's location sharing attitudes and found out that more complex privacy settings encourage people to share more. Another study revealed that people prefer to use combinations of simple privacy mechanisms, instead of a single one that does not meet all privacy needs (Burghardt et al., 2009). With regards to the level of concern towards location sharing, different studies have different outcomes. Barkhuus et al. (2008) observed that their participants had little concern over privacy when using an experimental location-based social network application. However, this study took place at the early days of location-based applications (for example the first iPhone was released in 2007) and it was based on a research application. On the other hand, a study that collected data from commercial location-based applications found that smartphone users are more concerned about their privacy than other Web users, especially in the context of online social networking (Li and Chen, 2010). In addition to this, a survey focusing on third party tracking, showed that 63% of their participants are concerned about that type of monitoring, whereas 50% are concerned about their location being monitored (Wills and Zeljkovic, 2011). These concerns can be

partially mitigated: Tsai et al. (2009) revealed that providing feedback to the users of a location sharing application reduces their privacy concerns, and further studies have shown that when users have a choice between different privacy settings but also between different location granularities, they feel more comfortable at sharing (Tang et al., 2012).

People's sharing attitudes towards location sharing vary and they are influenced by a number of factors. As explained in the previous chapter, several studies have shown that often privacy behaviours differ based on demographics. For instance, a survey focused on Brighkite [2] users (a commercial location-based social network) showed that factors like age, gender, mobility and geographic area influence users privacy concerns (Li and Chen, 2010). A comparative study between USA and China also highlighted differences (and similarities) between privacy preferences in location sharing between people from the two countries and also gender differences (Lin et al., 2013). It also appears that the preferences of the users are different depending on the purpose of location sharing; whether it is in exchange for a service or for social purposes (Tang et al., 2010). In an experiment involving a location-based application, people made decisions based on the perceived privacy and benefits from the available options, a result that is in line with the theory of the privacy trade-off (Knijnenburg and Kobsa, 2013b). A study in Locaccino (a location-based application developed by Carnegie Mellon University) showed that participants felt more comfortable sharing their location when in places with higher entropy (Toch et al., 2010). This also implies that participants did go through an evaluation of the pros and cons of sharing their location at locations with different entropy. It should be noted that location entropy is a measure of the diversity of people who visit a given location (Cranshaw et al., 2010).

4.1.2 The two roles of context in location sharing

This section aims to analyse the role of context in location sharing by looking at this concept from two different perspectives; a technical one that highlights the relationship between context and location data, and a social one that addresses the social context where a privacy decision takes place.

4.1.2.1 Context as a technical concept

Location information is part of a person's physical context (Duckham and Kulik, 2006). As a result, through location information other contextual information that refers to an individual may be inferred. Anonymisation techniques are a common means of securing user identity. However, latest research indicates that they are not sufficient at preserving the security of the data. An example that shows the potential of location data in de-anonymisation is the experiment conducted by de Montjoye et al. (2013). They

[2] brightkite.com

showed that with a dataset of location data they could uniquely identify 95% of the people in a large *anonymised* data set (approximately 1.5M users of a mobile phone operator). Several studies take advantage of the vast amount of location data that is now available by studying human mobility patterns (e.g. Noulas et al., 2012; Song et al., 2010; Cho et al., 2011). Song et al. (2010) managed to predict 93% of human mobility by studying people's mobility patterns and the entropy of their locations in an anonymised dataset. This outcome makes sense if we take into consideration that people travel in repeated patterns of mobility, therefore we can infer their future movements. Similar conclusions were reached by Cho et al. (2011), who studied three separate datasets and found common mobility patterns. Another study showed that social ties can be inferred by co-located photos uploaded in Flickr (Crandall et al., 2010). A similar example is Flap, a system that combines user location, content of messages and patterns in friendship formation in a large dataset to infer social ties (Sadilek et al., 2012). Flap also makes predictions about future locations based on their friends' locations. An interesting observation, stemming from this study is that other people can easily undermine the privacy of the individual. In the case of Flap the third parties were friends of the individual, however they can be any people connected to the same access point (e.g. in public hotspots) (Vratonjic et al., 2013). From these experiments it is apparent that the large datasets that Web applications currently offer an exceptional opportunity for inference mechanisms.

Other data that can be inferred through the publication of location data include people's activities, real-time emotional and physiological status (Riboni et al., 2009), and co-location (i.e. the presence of other people in the same location). Another example of a location-based inference refers to location entropy, which was described earlier. The entropy of the locations a person visits can indicate the number of social ties a person has within a network (Cranshaw et al., 2010). An interesting example is the application developed by Madan et al. (2010) that aims to predict people's health status not based on a health diagnosis but using sensors collecting location and communication-related data instead. A number of data such as co-location, entropy of interactions with others, and time of interactions helped researchers identify behavioural changes due to health issues. This means that not only can location generate inferences on other types of data, but when aggregated with other available information (e.g. metadata) even more powerful inferences can take place. An important aspect that needs to be taken into consideration when it comes to data inferences is the existence of historical data, due to the persistence of online data, which can help generate more inferences, such as predicting future user locations (Ruiz Vicente et al., 2011).

This plethora of contextual information that can be inferred through people's location in combination with the increase of applications that ask for people's location pose massive concerns about people's privacy. Scholars from various backgrounds have discussed the role of context in privacy management. Dourish (2004) distinguishes two approaches

towards the understanding of context in research; a technical approach, commonly used in Ubiquitous Computing and Human-Computer Interaction and an approach drawn from social science. In Ubiquitous Computing context refers to "the location of use, the collection of nearby people, hosts, and accessible devices as well as changes to these aspects over time" (Schilit et al., 1994). The matters we have discussed so far focus on the technical aspects of context, the next section explores its social aspects.

4.1.2.2 Context as a social concept

It is suggested by several scholars that context is a more complex concept than it is often suggested in research. Dourish (2004) describes it as "an emergent property of occasions of interaction, rather than a stable, objective set of features that externally characterise activity". He argues that research — in particular Human-Computer Interaction studies — should treat context as an interactional concept focusing on questions such as "how and why, in the course of their interactions, do people achieve and maintain a mutual understanding of the context for their actions?". In that sense, context is not static, but rather a dynamic concept generated and related to a specific activity. Similarly, Mancini et al. (2009) frame context in mobile privacy as place instead of space — space is a term more commonly used in Ubiquitous Computing. Space includes parameters such as GPS location, time, activities etc., whereas place include more subjective parameters related to social interactions and relationships.

In her popular book *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Nissenbaum (2010) describes social context as a structured social setting with dynamic characteristics and introduces the theory of *contextual integrity*. This theory offers a particularly useful means of understanding privacy expectations in a specific context. Contextual integrity refers to the desirable means of information release and it can be preserved when information appears only in an appropriate context and when its distribution is in line with the contextual norms of information flow. This theory describes accurately how privacy depends on context. As Palen and Dourish (2003) point out, the boundaries of privacy move dynamically as the context changes. Key aspects of the context-relative informational norms are the context, the actors (i.e. senders and recipients of information, and information subjects), the attributes (i.e. information types), and the transmission principles (i.e. constraints on the flow of information from one party to another in a specific context, e.g. information confidentiality between all actors).

Contextual integrity is a rather challenging task in today's digital world, because the contextual norms can easily be ignored by online applications (Barkhuus, 2012). The transmission principles are often disregarded, the contexts are much more diverse and complex than offline and there are also numerous issues concerning the actors taking part in information sharing. For instance, the audiences who have access (or may have

in the future) to people's online content are often different to the ones perceived by the individual when making a privacy decision. These audiences may include people outside of someone's network of friends in the case of a social network application, employers and colleagues, but they can also be third party applications. This issue was also highlighted by a study whose participants had negative reactions once they realised that their data was leaked and tracked by the applications they were using in their mobile phones — in fact they found it "creepy" (Shklovski et al., 2014).

Several studies have employed Nissenbaum's theory aiming to point out major privacy issues triggered by the use of popular web applications. Shi et al. (2013) provided a practical example of the key aspects of the theory through a qualitative study of friendship pages on Facebook (i.e. pages that show all the interactions between two Facebook friends, such as their wall posts, common tagged photos, comments they share etc.). Hull et al. (2010) analysed Facebook Applications (i.e. applications developed by third parties) and NewsFeed (i.e. a constantly updating list of stories from Facebook users) using Nissenbaum's theory and suggested design changes in order to make information flows more transparent to the users. Zimmer (2008) used contextual integrity in an attempt to illustrate how Google has changed personal information flows by collecting digital dossiers over its users.

4.1.2.3 Combining the two approaches

Taking both approaches towards context into account, we can assume that context in location sharing has a double role. First, it has a technical role, meaning that it refers to all the information than can be inferred through someone's location. Secondly, it has a social role that refers to the dynamic social settings under which information is shared.

The fact that privacy decisions do not only affect location data, but also other contextual data raises even more privacy issues. A study that analysed three different datasets showed that for different types of information people have varying degrees of disclosure (Knijnenburg et al., 2013). This leads to the acknowledgement that privacy decisions are *multidimensional*. They are dependent on a set of parameters, such as the given context of the decision, other contextual information that can potentially be inferred (time, duration, activity and so on), but also on the type of information released (in our case this is location, however numerous other information can be inferred). Apart from the multidimensionality of privacy decisions, it is of prime importance to stress that data is not any more ephemeral; the Web has converted contextual information into a permanent, globally and also easily accessed type of information. Grudin (2001) points out that the ability to control how information about oneself is interpreted in different contexts is either limited or in some cases completely absent. In the case of location data, when people trade their location with a system to what extent are they aware of the actual affordances of their location data?

Online data is persistent and privacy decisions are multidimensional, and it is rather easy to use location data to infer other types of contextual data. This implies that systems have now the ability to infer new information about people (possibly without their knowledge or consent) although in many cases people themselves would potentially not wish to disclose this information. Taking Nissenbaum's theory of contextual integrity into account, most inferences (if not all) that take place based on location data constitute privacy violations, since location data is used in ways outside of the contextual norms of the original transaction (unless the purpose of the information was clearly stated during the trade-off of the information). All these issues stress the fact that as technology advances, privacy becomes more complex and more precarious to handle. In the following section we will go through a survey of technical systems aiming to uncover the issues that are raised by location sharing decisions in a number of different systems.

4.2 Analysing location data and privacy

This section presents a structured analysis of how a set of technical systems deal with location data. It aims to understand the contextual elements of location privacy as described previously. Based on the background literature it appears that location privacy has a number of characteristics that mark it out from other privacy issues:

- Contextual data inferred through location includes information that can support surveillance, like tracking individuals and their activities.
- Location data deals with an individual's real-time location.
  The Web is a place where people can easily mask themselves; as the famous cartoon from the New Yorker says "on the Internet nobody knows you're a dog". Location data, however does not promote anonymity, as it enables systems to locate people explicitly in real-time — also argued by Cooper et al. (2010). A study, for example, showed that human mobility traces are rather unique by identifying 95% of the individuals in an anonymised dataset of location data (de Montjoye et al., 2013).
- Location sharing is not mainly employed by individuals as a means of persona building, as location data is usually published in exchange for a service (such as directions, new information). However, as this thesis later shows, in the case of social network sites individuals may share their location for impression management purposes.

In the previous section we presented context from a technical but also a social perspective. On the technical side we pointed out a number of contextual information that can be inferred through the exposure of location data. The purpose of this study is to investigate in more detail the relationship between location data and context from a

technical viewpoint. Given that at least a number of different contextual information can be inferred we aim to take a deeper look at the inferences that systems make based on location data. With that in mind, we develop a framework for analysis of location and its inferred data. Following that, we use the framework in practice in order to analyse a sample of technical systems presented in recent literature. This can potentially provide us with a clear idea of the actual affordances of location data.

4.2.1 Methodology

The foundation for the development of the framework was the analysis of technical systems that deal with location data. The first step was the selection of the systems that would be included in the analysis. Bearing in mind that the aim was to address the scope of location privacy we analysed all the systems presented in three years (2008, 2009, 2010) of ACM Ubiquitous Computing and ACM Mobile Human Computer Interaction conferences [3]. The list of papers used in the analysis can be found in Appendix A. We selected papers from these conferences, because they are premier conferences in Ubiquitous and Mobile Computing. As a consequence they are regarded as good predictors of future trends in this field.

The methodology we used includes a number of steps. Initially, any systems that deal with location data were selected for the analysis. Following that, a second selection process took place based on a set of criteria:

- The paper contains a commercial or research system.
- The paper focuses on location and contextual data.
- The exposed data are retrieved from hands-on experience with the system (either in the context of an experiment or real life usage).
- The exposed data refer to people.
- The paper is either a full or short paper of the conference.

Based on those criteria 32 research papers were selected, and 32 different systems identified. Following that step, we identified the location and contextual data exposed in each system based on a set of characteristics. In other words, a data category was selected for further analysis if it fulfilled at least one of the following characteristics:

- Explicitly discussed location data.
- Explicitly discussed contextual data.

[3] This analysis took place in 2011; hence the selected years of the systems

- Only data categories that were explicitly discussed in a paper were included in the analysis. Data categories that were not explicitly discussed, but could be inferred from the context were not included.

We use the term *data category* as a means of grouping together different types of data that refer to the same topic. In that sense, data categories spanned from location data (e.g. GPS coordinates, name of a city) to co-location and health data (e.g. user's health status). After the selection of data categories in all the papers of the selected sample, we analysed each data category based on a set of data properties, which are described in the following section. Finally, we conducted a numerical analysis of all the data categories in the selected papers, followed by an analysis of the numerical results. 164 data categories were identified in all the systems.

4.2.2 Data properties

This study aimed to analyse in depth the inferences that systems can make based on people's location data. We wished to answer questions that would help us understand deeper these inference mechanisms. These are presented in Table 4.1.

Question
1. How complex is the inference mechanism?
2. Is this information linked to a specific individual?
3. Has the individual given their consent?
4. How good is the quality of this information?
5. Who has access to this information?
6. Who is the source of this information?

TABLE 4.1: Questions addressed.

Based on these questions we developed a set of properties that aim to provide a deeper insight into that data and highlight the implications of exposing location and contextual information. Furthermore, these properties highlight the richness of the analysed information and the possibilities they have for profiling. An initial set of properties was defined based on the background literature and later refined with a small set of research papers retrieved from the Proceedings of the Mobile Human Computer Interaction conferences. In other words, the set of properties was used to analyse a set of data from a test sample of systems. In that way, this initial analysis verified the selected properties. These properties are useful as they manage to show not only what data can be inferred and aggregated, but they also address all the questions included in Table 4.1. Table 4.2

contains all the properties that were used in the analysis, and they are explained in the following subsections.

Data Degree	**Personally Identifiable Data**	**User Consent**	**Data Quality**	**Data Access**	**Data Source**
1st Degree	Directly	Explicit	Accurate	User	User
2nd Degree	Indirectly	Implicit	Complete	User Friend	System
3rd Degree	Heuristically		Timely	3rd Party	User Friend
	Non Identifiable			Everyone	3rd Party

TABLE 4.2: Data Properties.

4.2.2.1 Data degree

Looking at the different systems presented in these conferences it quickly became evident that some systems made simple inferences based on the users' location, whereas others used more sophisticated inference mechanisms. For example, Herbst et al. (2008) developed a mobile mixed reality game where the location of each participant was used to examine how close they were at a point of interest, therefore the inference was rather simple to make. On the other hand, Cranshaw et al. (2010) collected user location data from a social network site in order to develop a model that predicted friendships between people. Evidently, this was a far more complex inference mechanism requiring an algorithm to infer the new information.

Inspired by the background literature, where the issue of data inferences based on location was raised; the first property used in the analysis looks at how complex is an inference. Therefore, we called it *data degree* (Zafeiropoulou et al., 2012). This property addresses Question 1 in Table 4.1. With that in mind, location and contextual data can be classified into different degrees of data based on the complexity of the inference that generated them.

- **1st degree of data.** It refers to data that are not inferred through the system but are explicitly provided. For instance, in a location-based application the users explicitly declare their geographical location.
- **2nd degree of data.** Data that are implicitly inferred, e.g. the co-location between two users.
- **3rd degree of data.** Data that require inferences with more complex heuristics based on 1st and 2nd degree data. This may require the retrieval of data from a range of users.

The concept behind this classification of data can be further explained through the following scenario.

Scenario: Finding Alice
Alice is a regular smartphone user and allows her phone to update her location through a location-based application on a daily basis.
Mary, a friend of Alice, also a smartphone user and has the exact same functionality set in her own phone.
A third party collects and stores the tracks of users of this specific application. As a consequence, it is aware of the movements of Alice and Mary. The application also identifies and calculates the number of co-locations between the users. If the number of co-locations between any two users is significant, it is inferred that these two people are socially related. Apparently, Alice and Mary are often in the same location. Consequently, it is inferred that these two users are socially connected.

Overall, this scenario demonstrates the potential inference of several contextual elements in practice:

- location
- co-location
- activity
- social tie
- geographical hotspots

The above-mentioned contextual elements can be classified into different degrees of data based on their inference complexity. **Location data** is explicitly declared (i.e. no inference is required) and consequently belongs to the 1st degree of data. The 2nd degree of data refers to data that are inferred from location data, such as **activity** and **co-location**. In addition to this, the inference of co-location information makes use of data from Alice and from another user who is known to Alice (in this case Mary's data). The 3rd degree of data makes use of more complex heuristics, such as making inferences by combining Alice's data with the data from thousands of other users of the application who are unknown to Alice. An example could be the identification of social ties of users based on the number of co-location data between pairs of users. In the above scenario, the **social tie** between Alice and Mary could be inferred in that way. Another example could be the identification of **geographical hotspots** based on the users' location tracking.

4.2.2.2 Personally identifiable data

The second property addresses an issue that is of paramount importance when it comes to people's privacy. This issue refers to the use of personally identifiable information (PII), which includes any piece of data that identifies uniquely a particular person. It addresses Question 2, which was presented in Table 4.1. The location data we discuss in our analysis relates strictly to a device rather than a person. In that sense location data can potentially be personally identifiable information, especially in cases where they are combined with other pieces of information.

The reason behind the use of this property is that we wished to distinguish the data that were linked to individuals from anonymous data.

- **Directly Identifiable Data.** An individual is explicitly related to a piece of information. For example, in the case that a user shares their real-time location with a social network application, that location data is considered as directly identifiable.
- **Indirectly Identifiable Data.** It can be easily inferred that an individual is related to a piece of information.
- **Heuristically Identifiable Data.** It can be heuristically inferred with some probability that an individual is related to a piece of information. For example, a location keyword (i.e. the semantic name used by people to describe a location) can be heuristically identifiable by combining a set of heuristics (e.g. Lin et al. (2010) used machine learning algorithms to associate users with location keywords).
- **Non Identifiable Data.** A piece of information is not related to any individual. For example, time-stamp information was regarded as non-identifiable information.

4.2.2.3 User consent

A question that was raised at the end of the previous section focused on the extent to which people are aware of the affordances of their data. This question was also included in Table 4.1. With that in mind we developed the property *user consent*.

This property places its focus on whether the individual is asked to provide their consent before their location data is retrieved or published. User consent may be given not only **explicitly** but also **implicitly**, in cases where the user is not directly asked to give out their data, but the data are published with their full knowledge and the user does not take any action against it. User consent is only legally required for data that are PII.

4.2.2.4 Data quality

Another property that plays a significant role in location data exposure is the quality of the data (Question 4 in Table 4.1). The better the quality of data, the more accurate the inferences made upon it will be and consequently greater the threat to privacy. Data quality was calculated based on a set of three different characteristics that are commonly used in data quality studies (Wang et al., 2008; Wang and Strong, 1996), which are accuracy, completeness and timeliness:

- **Accurate Data.** The data is precise and objective.
- **Complete Data.** The data is complete in the sense that no values are missing from it or there is nothing to be added to it.
- **Timely Data.** The data is current and not out-of-date.

4.2.2.5 Data access

As part of our investigation into the different types of data that are inferred based on location, we also wished to identify who had access to that data (Question 5 in Table 4.1). As shown in Figure 4.1 there are a number of different entities who may have access to the data. In addition to this, they might have different types of access (read/edit/disseminate). It is assumed that the system has always access to the data. The sample is adequately described by a hierarchy, as shown in the figure, but of course it may be that a more complex structure is appropriate for a wider sample — for example a system might provide access to the data to itself and third party systems, but not to the user or their contacts.

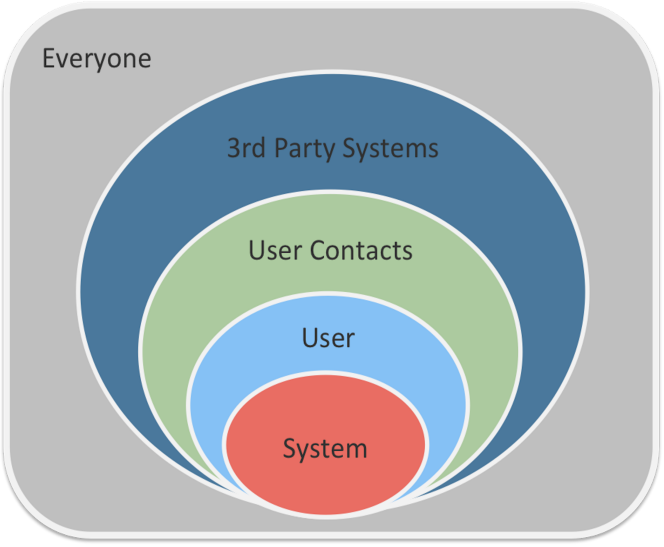


FIGURE 4.1: Who has access to data.

4.2.2.6 Data source

The final property dealt with the origin of the data and addressed Question 6 in Table 4.1. There can be a number of different sources of data, such as the user, the system or even friends of the user and 3rd parties.

4.2.3 Analysis

After selecting the data categories that appeared in all the selected systems, each of the systems was analysed based on the properties described above and displayed at Table 4.2. The next step was the analysis of the sample with the purpose to shed light on how these systems handle people's data.

First, we identified the different types of systems that make use of location data. As Figure 4.2 shows a number of different types of systems use location data.

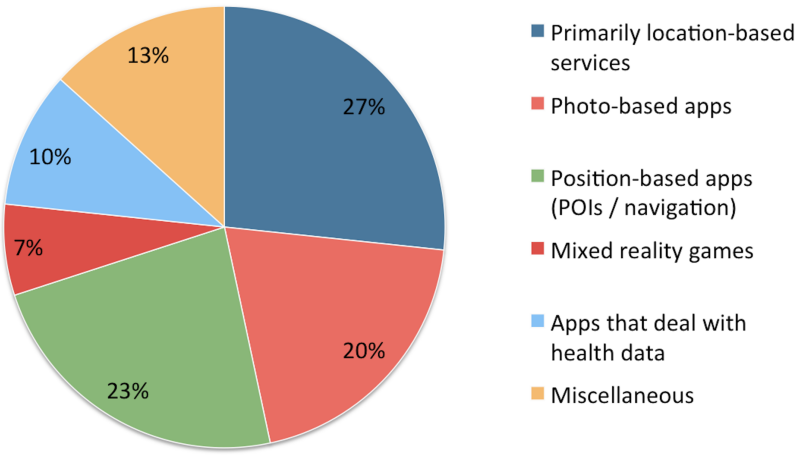


FIGURE 4.2: Types of systems identified in the analysis.

Half of the systems do make 3rd degree inferences, specifically 15 out of 30 (Figure 4.3). We also identified systems that make 3rd degree inferences where the inferred data go beyond the context of the person's location, i.e. where the inferred data have no semantic relation with the 1st degree location data. Such an example is the inference of social ties between users based on the number of times they were co-located. Out of the 15 systems that make 3rd degree inferences, 5 use location data to make inferences beyond location.

We undertook an analysis of the different properties along the dimension of the *data degree* property. For each of the relevant properties, we analysed our sample to see whether the systems treated inferred data (2nd and 3rd degree data) differently from

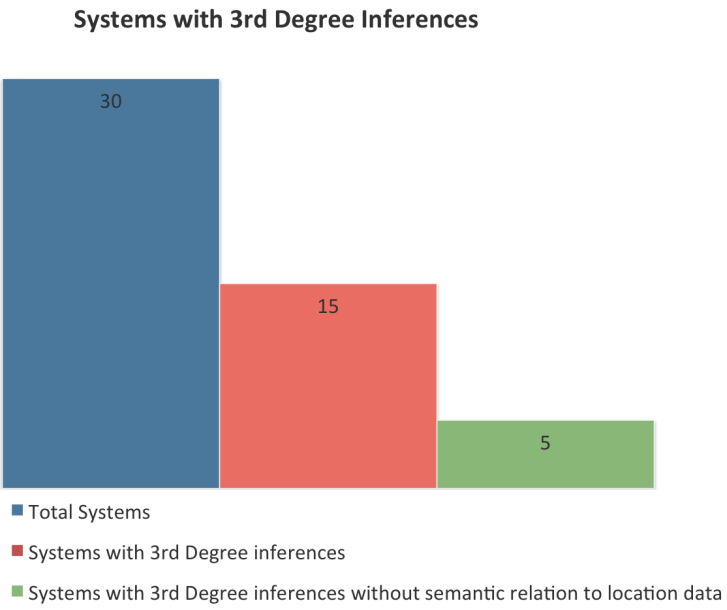


FIGURE 4.3: Systems that make 3rd degree inferences.

the data that was released by the users (1st degree data). The detailed analysis table can be found in Appendix A.

4.2.3.1 Personally identifiable data

Although there was considerable variation based on data degree, the majority of the data in all cases was personally identifiable (Table 4.3). That means that most of the information could easily be associated with a specific individual.

	1st degree	2nd degree	3rd degree
Directly	**67%**	**56%**	**43%**
Indirectly	11%	12%	7%
Heuristically	3%	6%	4%
Non Identifiable	18%	24%	39%

TABLE 4.3: Degree-based Analysis of Personally Identifiable Data.

4.2.3.2 User consent

With regards to 1st degree data most systems expected users themselves to expose their data (e.g. user location), so it was taken for granted that the user consent was given. However, when it came to 2nd and 3rd degree data there was not sufficient information to suggest that the consent of the user was requested (Table 4.4).

	1st degree	2nd degree	3rd degree
Explicit	**69%**	21%	18%
Implicit	15%	21%	18%
No Info	16%	**44%**	**39%**

TABLE 4.4: Degree-based Analysis of User Consent.

4.2.3.3 Data quality

Most of the systems were provided with high quality of 1st degree data in terms of completeness, timeliness and accuracy, especially because the data of that degree are user-generated. However, in many cases there was not sufficient information with regards to the quality of 2nd and 3rd degree data (Table 4.5).

	1st degree	2nd degree	3rd degree
Good Quality	**55%**	18%	7%
Low Quality	9%	24%	43%
No Info	36%	**59%**	**50%**

TABLE 4.5: Degree-based Analysis of Data Quality.

4.2.3.4 Data access

As Table 4.6 shows, regardless the degree of the data the majority of the data in these systems were available to the user who they refer to. Nevertheless, in most cases the access rights of the user were not clear in the papers. It is worth pointing out that in most of these systems there was no clear indication about 3rd party systems involved.

	System	User	User Contact	Everyone	Unknown
1st degree	4%	**63%**	22%	5%	6%
2nd degree	15%	**67%**			18%
3rd degree	36%	**46%**			18%

TABLE 4.6: Degree-based Analysis of Data Access.

4.2.3.5 Data source

As expected 1st degree data were mostly user-generated, whereas 2nd and 3rd degree were generated by the system (Table 4.7).

	System	User	Unknown
1st degree	47%	**53%**	
2nd degree	**79%**	21%	
3rd degree	**93%**	3%	4%

TABLE 4.7: Degree-based Analysis of Data Sources.

4.2.3.6 The role of location data

Finally, the analysis unveiled the role of location on the type of data that the analysed systems publish. We looked at the data categories in each system and given the level of inferences that took place based on location we characterised the role of location as *primary*, *secondary*, or *minor*. In location-based systems location affects primarily the type of location published, however location was not a key piece of data in all systems.

As Figure 4.4 shows, in more than half of the systems location plays a primary role on the type of data that are published about a user. 25% of the systems are affected by location data but not primarily, whereas 16% use location data only as metadata.

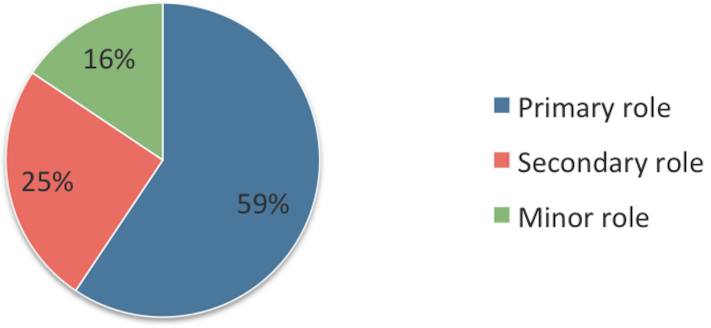


FIGURE 4.4: The Role of Location in the analysed systems.

4.2.4 Discussion

The results described in the previous section highlight the power of location data as a starting point for aggregating and inferring data. For instance, Figure 4.3 illustrates the level of inferences that take place in the analysed systems. According to the Figure one third of the inferred data are inferences on an entirely new piece of information about the user (e.g. social ties) and therefore have no semantic relation to location data. Figure 4.4 confirms the role of location data as a catalyst for linking data across the Web.

As the results revealed, the majority of the data in the analysed systems could be associated with a specific individual (Table 4.3). Although in the majority of the systems the 1st degree data were exposed with the individual's explicit consent, there was not sufficient information to suggest that the consent of the users was requested before making 2nd and 3rd degree inferences. This implies that only a minority of the systems are explicitly concerned about privacy. The majority of the systems do not take privacy into account and do not propose privacy mechanisms. As a consequence, it appears that as the degree of the inferences moves from 1st to 3rd degree, the individual may lose control over their data. According to Nissenbaum's theory of contextual integrity, which was described in the previous section, many inferences that take place constitute privacy violations, since in several systems location data is used in ways outside of the contextual norms of the original transaction.

Taking into account these observations, it can be assumed that the exposure of location data on the Web may cause a number of privacy related risks, as it enables a number of inferences that include personally identifiable data. This framework allows a more targeted investigation of the relations between the complex issues of consent, inference and access. The framework addressed a number of questions that are related to the topic such as "what level of complexity does a specific inference require", "who has access to this data" with the aim to assess these inferences. The framework can potentially be employed to understand the potential privacy risks of large datasets containing inferred data. Apart from the questions we addressed, there are other questions that can be raised with relation to data inferences, such as "what is the computational cost of 3rd degree inferences" or "how important is the functionality of the inferred information". However, in our analysis we decided to focus on a set of basic questions related to data inferences that would allow us to explore the privacy-related issues that may arise.

At this point, the limitations of this analysis should also be pointed out. The majority of the systems were research systems and not commercial systems. The data were collected in many cases in the context of an experiment instead of real usage of the systems. In addition to this, in many cases there was not sufficient or clear information to suggest the quality of the data, the consent of the user or even whether the system took any actions to anonymise the collected data.

This framework is intended to offer a method for analysing location data. We used it in our attempt to unravel the complexity of location privacy. Above all, this framework stresses the lack of awareness when inferences are made based on people's location data; people are not aware of the actual affordances of their location data.

4.3 A Distance Model of Belief, Behaviour and Affordance

Based on the findings of the analysis, which was described in the previous section, we developed a theoretical model called the Distance Model of Belief, Behaviour and Affordance (DMBBA). The model demonstrates the distance between people's beliefs and their actual behaviour with regards to location privacy. Although this model is based on the privacy paradox it adds a new aspect to the matter, which is the way systems actually use people's data and its distance from people's beliefs with regards to the privacy of their data. For instance, the analysis highlighted that systems may easily infer whether two different users are co-located (Ruiz Vicente et al., 2011) as well as whether they are friends in an online social network (Cranshaw et al., 2010). Systems may also infer with a certain probability, whether a user is in a good state of health (Madan et al., 2010). A question of significant importance emerges from these observations: To what extent are users aware of these potential inferences when they publish their location online?

Figure 4.5 illustrates the model as a diagram. The figure shows people's perspectives with regards to the handling of their data, but also systems' perspectives with regards to the ways through which they handle people's data. The diagram consists of three nodes: the left node refers to people's beliefs with regards to the privacy of their data, the central node refers to people's actual disclosure behaviours and the right node refers to the way systems use people's data. The left node is linked with the other two nodes with two arrows. Each of these represents the distance between people's beliefs and reality (in the case of the central node actual disclosure behaviour and in the case of the right node the actual affordances of data). The first distance is called the "Belief-Behaviour" distance, whereas the second one is called "Belief-Affordance distance". The second distance is evidently greater than the first one. The reason behind this is that the affordances of people's data are actually far greater than they are even aware of. In that sense, the model aims to show that the "Belief-Affordance" distance is actually more serious than the paradox indicates. The notion of *distance* in this model is not used a quantitative measure, but as a means to emphasise the contrast between people's beliefs and reality.

4.3.1 The Belief-Behaviour distance

One of the focal points of this thesis is the Belief-Behaviour distance, which is the first distance shown in Figure 4.5. It represents the distance between people's attitudes towards location privacy on the Web and their actual disclosure behaviour. According to the privacy paradox there is a discrepancy between people's privacy intentions and their actual behaviour. Several research projects have provided evidence regarding a paradoxical behaviour of people when it comes to privacy, which were presented in

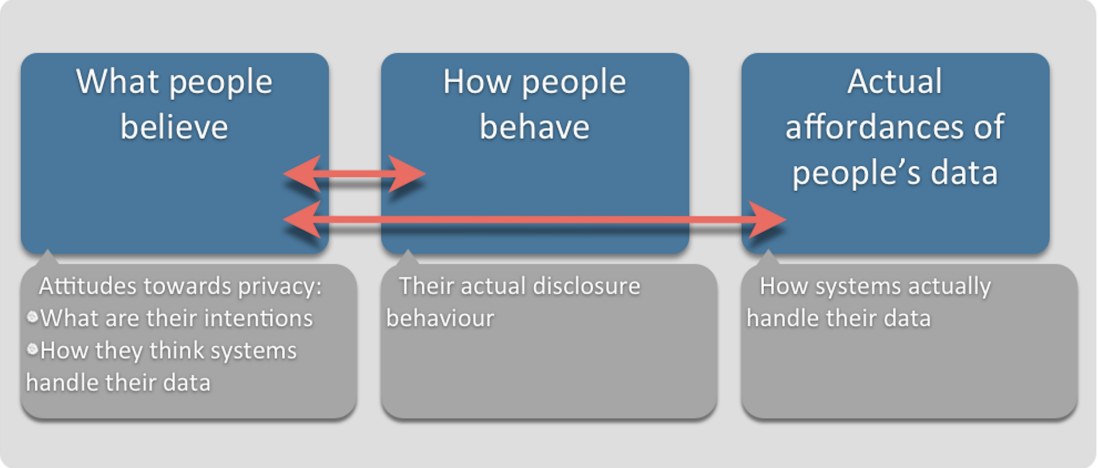


FIGURE 4.5: A Distance Model of Belief, Behaviour and Affordance (DMBBA).

detail in the Chapter 2, thus justifying the existence of the Belief-Behaviour distance. This thesis investigates whether the paradox applies to location data as well. We will explore this part of the model further in the following chapter.

The Belief-Behaviour distance is also consistent with the theory of cognitive dissonance (Festinger, 1957), which was described in Chapter 2. Both of them highlight the issue of conflicting beliefs and behaviours. As we explained in that chapter, Festinger's theory suggests that people have a strong inner motivation to find a balance between contradicting attitudes and beliefs in order to be consistent internally, and hence avoid dissonance. In the case of privacy decision-making people have certain attitudes towards privacy, however their actual disclosure behaviour is often different from their stated attitudes. According to cognitive dissonance, people try to rationalise their disclosure behaviour to avoid the tension caused by this discrepancy (e.g. the likelihood that their privacy could be seriously violated is rather small). In that sense, the distance presented on the left side of Figure 4.5 reflects the cognitive dissonance that arises from the conflict we often observe between privacy attitudes and privacy behaviour.

4.3.2 The Belief-Affordance distance

Apart from the Belief-Behaviour distance, DMBBA explores people's lack of awareness regarding the actual affordances of their data. This is displayed in Figure 4.5 through the distance between *what people believe* and *the actual affordances of their data.*

Usually applications make use of people's location data explicitly with user consent, however they are capable of further manipulating people's data without the knowledge of the users. The analysis described earlier highlighted the fact that simply by releasing their location people empower applications to make a number of probabilistic inferences. When people allow applications to access their location they should be aware of all the

benefits but also the costs of each privacy trade-off. However, this is not the case in reality. As we discussed in Chapter 2, people's privacy decisions are restrained by limited information about the outcomes of their trade-offs (Acquisti and Grossklags, 2005). Bounded rationality — the concept arising from limited information in decision making — in the case of the Belief-Affordance distance causes a technological dissonance, as people are unaware of the actual affordances of their data (compared to the Belief-Behaviour distance that reflected a cognitive dissonance between people's beliefs and behaviour).

This distance highlights the importance of developing transparent systems. Lately, researchers have begun developing methods and tools that can potentially assist people in this matter (e.g. Krishnamurthy, 2013; Malandrino et al., 2013). It should be clarified that transparency requires not only technical solutions but also policy frameworks. As soon as people release their data to an application, they lose control over their data, as the results of the analysis of the technical systems indicate. In tandem with this, as we pointed out in the previous chapter (Chapter 3) the affordances of people's data make the issue of transparency more challenging. However, in a scenario where this issue was appropriately addressed, and as a result the data manipulation by the systems was completely transparent, the second distance of Figure 4.5 (between people's beliefs and the actual affordances of their data) would not exist.

The DMBBA model is not a complete description of disclosure decisions, but it does however integrate all the points made in the previous chapter with regards to the privacy trade-off with the privacy-related issues of location data, as well as provide a path to develop this research further. By designing the DMBBA and raising the related issues, this thesis aims to understand in more depth the nature of the two distances as described in this model (Figure 4.5).

4.4 Conclusion

The first part of this chapter was a review of current research on location privacy on the Social Web. The review also explored the role of context in location sharing. First, due to the fact that location is part of someone's physical context, it has the potential to infer other contextual information. This is further studied in the second part of the chapter. Second, context is also a social property related to the dynamic settings under which a privacy decision takes place.

In the second part we investigated the scope of location and contextual privacy. To that purpose we developed a framework for analysis of location and contextual data. The framework was used to analyse location data exposed in a sample of 32 technical systems. The outcomes of this analysis confirm the inferential power of location data. It also raises the issue that several inferences produce non-contextual data (such as

someone's health status), which may cause further privacy concerns. We argue that this important characteristic of location data has implications for privacy management systems, as they can potentially fail at being transparent with their users regarding the actual affordances of their data. As a consequence, people release their data online with incomplete information regarding the actual usage of their data.

Bearing that in mind, we developed a theoretical model that connects this issue with the privacy trade-off and cognitive dissonance, called DMBBA. The purpose of this model was to uncover the issue that is raised from the existence of the two "distances", people's beliefs regarding online privacy and their potential dissonance from the affordances of their data. The following chapter focuses on the "Belief-Behaviour" distance of the model, as it aims to investigate the paradoxical nature of privacy decisions and unfold the reasoning behind these decisions.

Chapter 5

Uncovering Location-based Disclosure Decisions

The framework presented in the previous chapter highlighted the primal role of location data in making probabilistic inferences about other types of information. This chapter takes a step further, as it attempts to focus on the perspective of the users. The research objective is to gain a deeper understanding of the way in which people trade their data in exchange for services, how they value their data in an abstract or objective sense, and how they justify their decisions during the trade-off, based on the theories described in Chapter 2. The methodology used comprises of the design and dissemination of a survey followed by a quantitative and a qualitative analysis of the results. All the stages of the study are described in detail in the following sections.

5.1 Methodology

We employed a survey in the form of an online questionnaire, in which participants were prompted to answer questions regarding their attitudes towards online privacy and their location sharing decisions. The survey was aimed at people who connect to the Web through their mobile devices (i.e. smartphones or tablets) — it can be found in Appendix B. It gained ethics approval by the University of Southampton Ethics Committee [1] (Ethics reference number: 1521).

The design of the survey included several steps: *a*) mapping the research objectives to survey questions to ensure coverage, *b*) designing real-life scenarios to elicit realistic behaviour, and *c*) validation through a small pilot study. Table 5.1 illustrates the first step, and more specifically it shows how the research questions were mapped to the different types of questions in the survey.

[1] www.ergo.soton.ac.uk

Research Question	Survey Sections
How do people perceive and value their location privacy in theory?	8 Likert-scale questions
How do people value their location data in practice during the privacy trade-off?	5 scenario-based questions using multiple-choice
To what extent do people act as agents and to what extent are they influenced by structures during the trade-off?	5 qualitative questions on justifying their choices (1 for each scenario question)

TABLE 5.1: Mapping the research questions to survey questions.

The first section of the survey consists of a set of demographic questions with the hope to explore potential differences in privacy attitudes based on demographics.

The second section of the survey included a set of scenario-based questions that aimed to shed light on the privacy trade-off (second row in Table 5.1). Participants were asked to choose from a set of Web-based applications the ones they use in practice (Wikipedia, Facebook, IMDb, Twitter, and Foursquare). Depending on the applications they chose, they were directed to different pages that contained a scenario-based question for each of the chosen applications. A scenario was presented to them and then they were asked to decide whether they would share their location in this context. They were prompted to decide between three choices: "Yes", "Maybe", or "No". For example, the following is the Facebook scenario that was used in the survey:
"Consider the following scenario. You are visiting a friend (who is also your Facebook friend) in another city. You are going to dinner in a very popular restaurant of that city. Would you post your location on your Facebook wall?"

The scenario-based questions were used to explore people's location sharing decisions, and they were designed in this way to attract as spontaneous answers as possible. A common challenge for empirical studies on privacy decision-making is to gather information regarding the actual privacy disclosure mechanisms that individuals employ. In that sense, survey methods cannot easily gather information about how people feel about privacy in practice (Mancini et al., 2009). This survey therefore studies individual's location privacy behaviours by placing people in a real life scenario (i.e. in context). In that sense, the questions were framed within scenarios from the participants' every day use of the Web. In addition to this, participants had also the opportunity to justify their answers to the scenario-based questions (in the next section of the survey) in a way that was concrete and situated in each scenario.

After they replied to all the scenario questions, participants were directed to the third section in which they were asked to justify their answers (third row in Table 5.1). The

justifications were open text fields, where participants could write free text. They allowed us to better explore the mechanism of the privacy trade-off, as well as the extent to which participants act as free agents in their privacy decision-making.

The final section of the survey contained the Likert-scale questions with regards to location privacy attitudes and people's privacy concerns (first row in Table 5.1). Up until this section, none of the questions were explicitly privacy related. This was done on purpose, as research has shown that the wording used in privacy-related surveys plays a significant role in the way participants answer; questions that contain privacy-related language have a strong effect on the way participants answer (Braunstein et al., 2011).

Prior to its dissemination we validated the survey with a pilot study, where a small number of test participants answered the survey questions. The aim was to uncover any ambiguities and ensure that the questionnaire was to be completed in a reasonable time (ten to fifteen minutes). During this time we also made minor revisions to the wording of the questions.

5.1.1 Analysis objectives

At this point we also planned the quantitative and qualitative analysis of the survey. With regards to the quantitative analysis, calculating the simple mean for the Likert-scale questions, could paint a picture of people's stated privacy attitudes. These could then be correlated with the responses to the scenario questions to see if stated privacy attitudes could predict contextualised privacy decisions. In addition to this, we planned a qualitative analysis of the justifications for each scenario to take place, in order to explore the factors behind these decisions. Coding of the data was based on a qualitative thematic analysis (Seale, 2004; Braun and Clarke, 2006) using NVivo 9.

In that sense, the main objectives of our analysis are the following:

- Statistical analysis of Likert-scale questions.
- Investigation of participants' answers to the scenario-based questions. This may provide insight into the relation of structuration and the privacy trade-off. To what extent people act autonomously as agents and to what extent are they affected by certain structures.
- Exploration of the privacy trade-off; why people decide to share or not their location data with these applications (how they justify their location sharing decisions).
- Investigation of the privacy paradox; whether people's answers to the scenario-based questions reflect their answers to the Likert-scale questions.

- Identification of potential gender differences in participant responses; do men share more online than women as previous research has often pointed out (as presented in Chapter 2).

5.2 Quantitative analysis

The survey was disseminated through online social media (Facebook, Twitter, and LinkedIn), but also through the mailing list of the research group at the University of Southampton. It remained open for 8 weeks and received 150 responses. This section presents all the results of the statistical analysis.

5.2.1 Demographics

Among the 150 participants, the majority were male (90 participants, 60% of total); however the number of female respondents was significant too (60 participants, 40% of total).

Table 5.2 illustrates the age groups of the respondents. The majority of the respondents were young people with ages between 18 to 34 years old.

Age group	Number	Percentage
18-25	39	26%
26-34	87	58%
35-43	16	10.7%
44-on	8	5.3%
Total	150	100%

TABLE 5.2: Age groups of participants.

Table 5.3 illustrates the countries that the majority of the participants came from. 56 participants came from the UK, 48 came from Greece — numbers that make sense if we take into account that the research was undertaken by a Greek student living in the UK — whereas the remaining 46 came from 25 other European, Asian, American and African countries (and are not displayed in the table).

Country	Number	Percentage
United Kingdom	56	37.3%
Greece	48	32%
India	6	4%
Spain	6	4%
United States	6	4%
Germany	3	2%
Saudi Arabia	3	2%

TABLE 5.3: Country of Origin.

5.2.2 Scenario-based questions

Before coming across the different scenarios, participants were prompted to choose what type of applications they use in practice:

- Wikipedia
- Social Networks (Facebook, Google+, etc.)
- Movies, Music and Event Planner (IMDB, Flixster, etc.)
- Microblogging Applications (Twitter, Weibo etc.)
- Location-based Social Networks (Foursquare, Gowalla, etc.)

Application Type	Number	Percentage
Wikipedia	79	52.7%
Social Networks	126	84%
Movies, Music and Event Planner	56	37.3%
Microblogging Applications	66	44%
Location-based Social Network Sites	18	12%

TABLE 5.4: Applications used by participants in their every day life.

Table 5.4 illustrates the number of respondents who use each of these types of applications. Social network sites are by far the most popular applications, as 126 respondents (out of 150 respondents overall) make use of them. Wikipedia is also very popular with 79 respondents out of 150 overall visiting it. 66 of the respondents use applications like

Twitter, whereas 56 of the respondents use applications such as IMDb. Interestingly, only 18 of the respondents use location-based social network sites (e.g. Foursquare).

Depending on which type of application they use, participants were directed to the corresponding scenarios and were asked whether they would share their location data. Table 5.5 illustrates participants' answers to the scenario-based questions.

Application	Yes	Maybe	No	Total Answers
Wikipedia	54	8	14	76
Facebook	19	33	69	121
IMDb	29	4	21	54
Twitter	20	13	32	65
Foursquare	6	3	9	18

TABLE 5.5: Answers to the scenario-based questions.

Figure 5.1 illustrates participants' answers from Table 5.5 as percentages in a stacked bar chart and highlights which of the applications the participants trusted more with their location data. Evidently, Wikipedia was the most trusted application, followed by IMDb. On the other side, Facebook was by far the less trusted application. Twitter and Foursquare users were almost equally divided between people who are negative about sharing their location on these applications and people who are either positive or thinking about sharing their location.

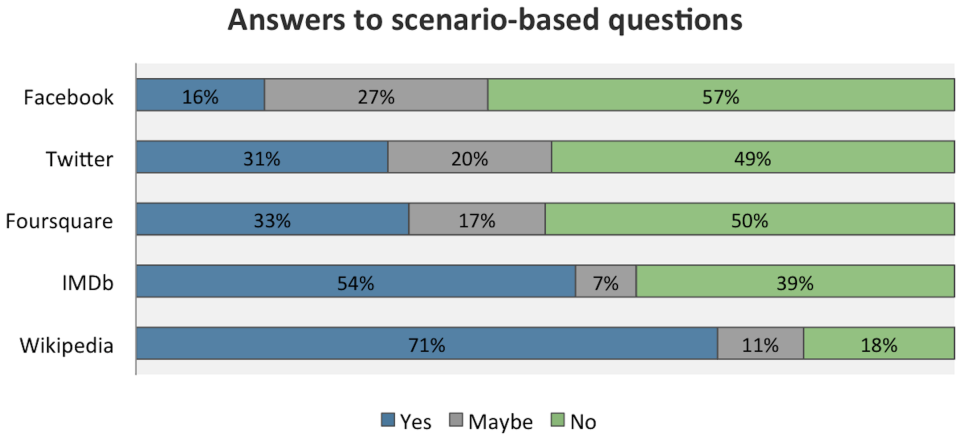


FIGURE 5.1: Percentages of answers to scenario-based questions.

5.2.3 Privacy attitudes

The last section of the survey includes theoretical questions regarding people's privacy perceptions using Likert-scale responses. 125 out of 150 participants completed the

survey till the end and replied to all the theoretical questions.

The following figures illustrate participants' answers to a set of questions regarding their attitudes towards online privacy and location sharing.

Figure 5.2 shows the respondents' replies to the following two questions that deal with the way people control their online privacy:

- **Question I.** I believe I am able to take the appropriate steps to control when and how my location is released online.
- **Question II.** When an application requests my location, I am fully aware of the reasons why.

Each column in Figure 5.2 represents the number of participants who chose the corresponding option.

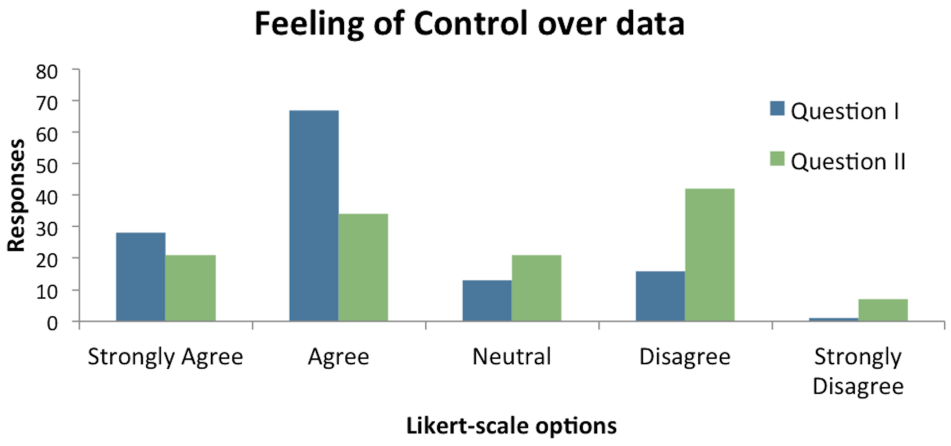


FIGURE 5.2: Responses to Question I and II.

Similarly, Figure 5.3 shows people's level of concern regarding their online privacy in general as well as location privacy. The questions they were asked are the following:

- **Question A.** How concerned are you about threats to your online privacy?
- **Question B.** How concerned are you about the fact that your location might be used for other purposes too?

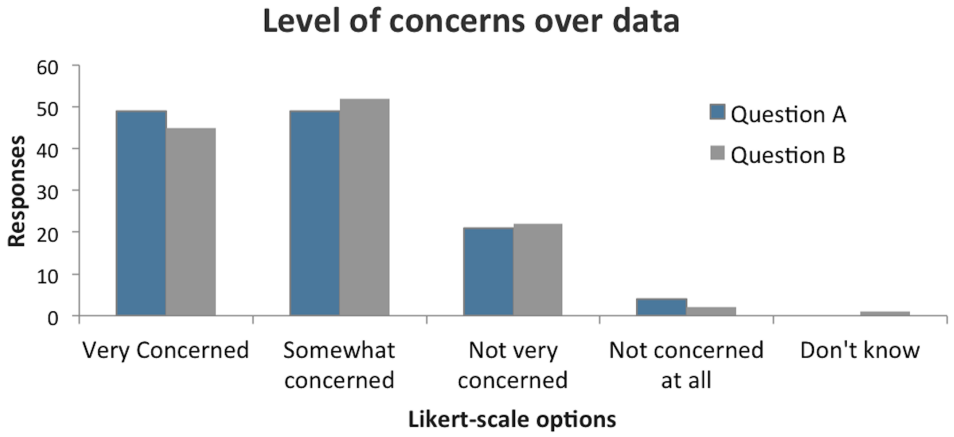


FIGURE 5.3: Participant concerns over their data.

As anticipated, the majority of the participants (approximately 80% for both questions) replied to these questions that they are either very concerned or somewhat concerned. More specifically, in Question A 40% of the participants answered they are very concerned about threats to their online privacy, 40% replied that they are somewhat concerned about their privacy, whereas only 17% answered that they are not very concerned. With regards to Question B the majority of the participants (43%) replied that they are somewhat concerned about threats to their location privacy and 37% answered that they are very concerned.

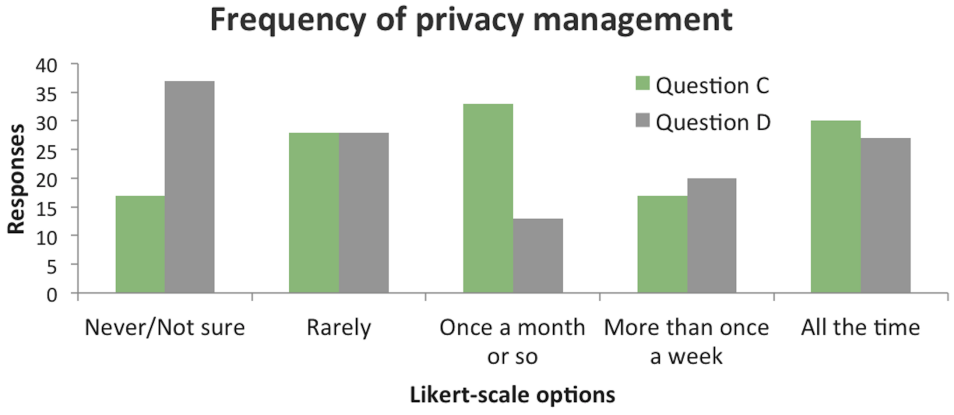


FIGURE 5.4: Participant answers regarding Questions C and D.

Figure 5.4 illustrates participant responses to questions that deal with the management of their privacy settings:

- **Question C.** Do you make use of the privacy settings offered by Web applications to control access to your data?

- **Question D.** In your mobile device do you ever have the location services setting on?

In Question C the answers were distributed among all the possible choices. Still, the majority (64%) stated that they use their privacy settings at least once a month. In Question D the answers were also distributed among all the choices. The largest response group (30%) answered that they are not sure or they never have the location settings on in their devices, but a significant minority (48%) had location services on at least once a month.

Figure 5.5 shows people's responses to questions that deal with their online location sharing attitudes:

- **Question E.** How often do you post your location in a social networking application (Facebook, Twitter etc.)?
- **Question F.** Do you ever allow an application to determine your current location?

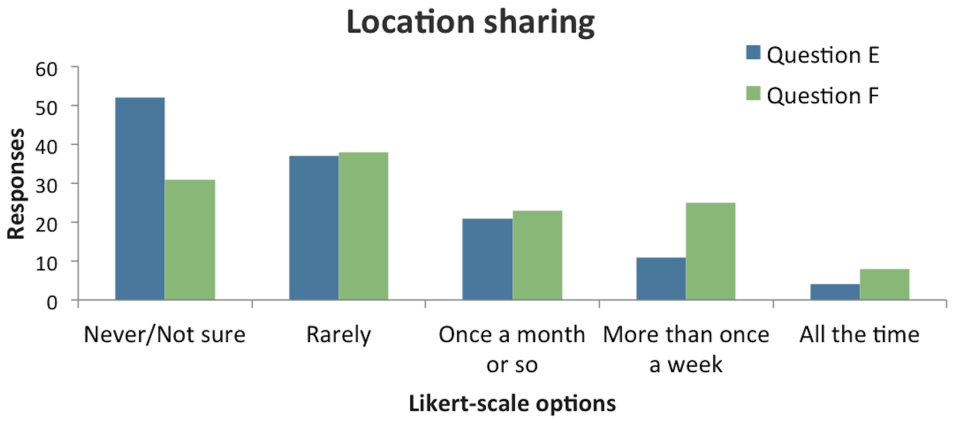


FIGURE 5.5: Participant answers regarding Questions E and F.

In Question E most participants answered "Never/Not sure", whereas in Question F all answers were distributed among all the options. In both questions only a very limited number of people (3% for social networks and 6% for apps) published their location all the time.

The final question was based on the outcomes of the study presented in the previous chapter, which stressed the primal role of location data in inferring other types of information. The question participants were asked was the following:

- **Question G.** Location-based websites may use your location to make assumptions about you. Please indicate how important it is for you to control these assumptions for each of the following types of information.

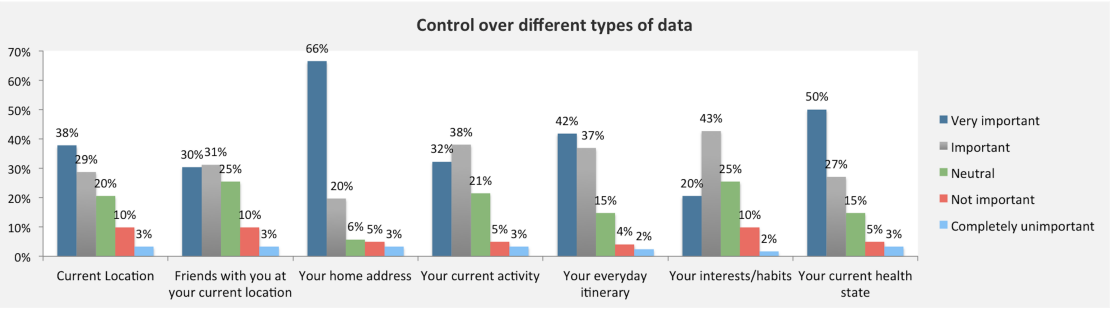


FIGURE 5.6: Importance of controlling inferences made by applications regarding certain types of information

For any given type of information the answers varied; with the exception of people's home address. The vast majority of participants (86%) found it important to control inferences based on their home address. However, as Figure 5.6 shows, for all types of information the majority of the responses were either "Very important" or "Important". In particular, 77% of participants found it important to be able to control inferences about their current health status, 63% about their interests, 79% about their every day itinerary, 70% about their current activity, 61% about their co-located friends and 67% inferences on their current location. These results show that regardless of their sharing decisions, people wish to be able to control potential data inferences and aggregations.

5.2.4 The privacy paradox

By looking at all the graphs of the previous section it is evident that in most questions participants' answers were highly dispersed between all the possible options. However, a number of observations were made:

- A majority of people are concerned about their privacy (80% responded concerned or somewhat concerned).
- A majority of people use privacy settings (64% responded that they use privacy settings at least once a month).
- A majority of people actively restrict access to their location (94% allow applications to access their location once a week or less, whereas 97% post location on social networking sites once a week or less).
- A majority of people find it important to be able to control potential inferences based on location on various types of information (77% on current health status, 63% on interests, 79% on every day itinerary, 70% on current activity, 86% on home address, 61% on co-located friends and 67% on current location).

These findings are therefore in line with previous surveys on privacy attitudes. Existing work on personal data also discusses the existence of a privacy paradox, i.e. a dichotomy

between attitudes towards privacy and actual disclosure behaviour. This study aimed to investigate whether the privacy paradox also applies to location data.

The relationship between the responses of the participants in Questions A and B (presented in the previous section) and their responses in the scenario-based questions, which were presented in section 5.2.2, could verify the existence of the paradox. To test this assumption, the different options in these two questions were transformed into numerical variables to represent *Concern*, with values "Very concerned" = 3, "Somewhat concerned" = 2, "Not very concerned" = 1, and "Not concerned at all" = 0. A new variable was also introduced named *WillingnessToShare*, which was the score of each participant's answer in the three most popular scenarios (Facebook, Wikipedia, and Twitter). The score was calculated based on the answers the participant gave in each scenario ("Yes" = 2, "Maybe" = 1, and "No" = 0) divided by the number of answers. For instance, if the participant answered "Yes" in the Facebook scenario, "Maybe" in the Wikipedia scenario and they did not answer the Twitter scenario (because they don't use Twitter) the *WillingnessToShare* in this case would be 1.50.

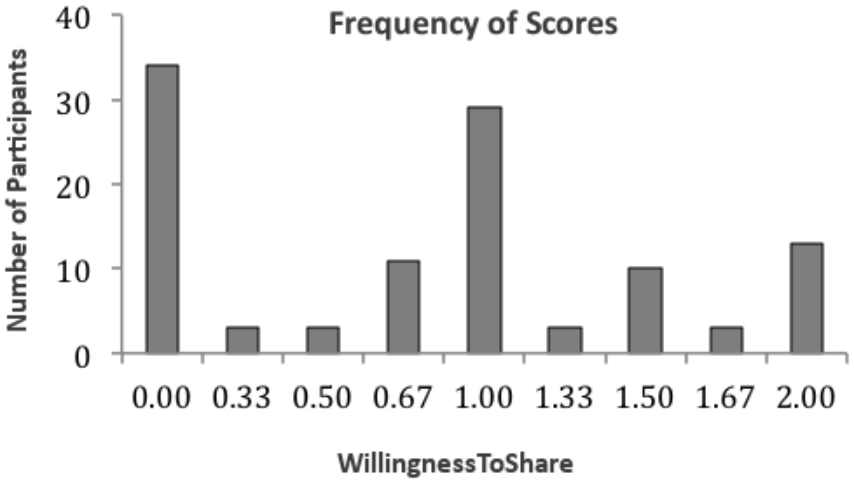


FIGURE 5.7: Participant score with regards to location sharing.

Figure 5.7 illustrates the frequency of different scores between the participants ranging from 0.00 to 2.00. 0.00 represents participants who answered "No" in all the scenarios, whereas 2.00 that refers to participants who answered "Yes" in all the scenarios.

Figure 5.8 and Figure 5.9 illustrate two scatter plots, where *WillingnessToShare* is in the horizontal axis and *Concern* (derived from Question A or B) is on the vertical axis. It is worth pointing out that some points in these plots represent a single response, whereas others represent many responses with the same values. A visual inspection of both plots suggests that there is no correlation between these variables. However, we wanted to test statistically the actual correlation between them. We executed the Kolmogorov-Smirnov and the Shapiro-Wilk tests for normality for all three variables — i.e. Question

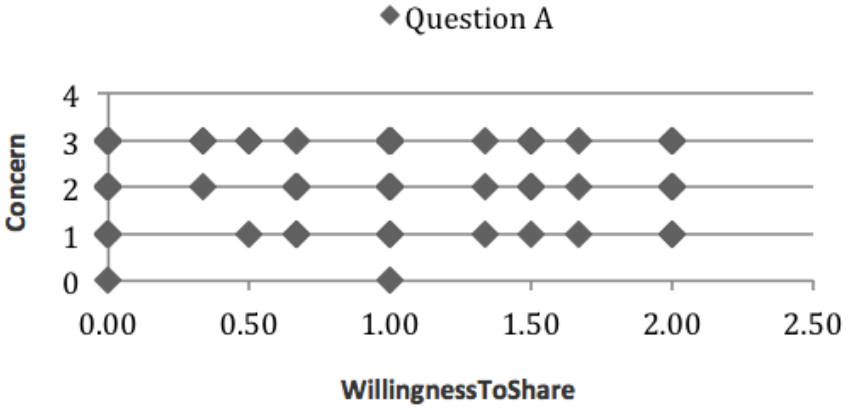


FIGURE 5.8: Plot between *WillingnessToShare* and *Concern* in Question A.

A, Question B, and *WillingnessToShare* — and discovered that in all cases the data is not normally distributed ($Sig. = 0.01$). For this reason we conducted both a Pearson's and a Spearman's correlation test (assuming that Spearman's would be less sensitive to outliers). The results of the correlation between Question A and Question B showed that there is a moderate correlation between them, which is statistically significant (Pearson's $r = 0.527$, $p = 0.001$ and Spearman's $r_s = 0.528$, $p = 0.001$). In other words, people who are concerned about privacy in general also tend to be concerned about the privacy of their location data. However, the results of the correlation between *WillingnessToShare* and *Concern* measured in Question A (Pearson's $r = 0.067$, $p = 0.492$ and Spearman's $r_s = 0.073$, $p = 0.453$), as well as in Question B (Pearson's $r = 0.103$, $p = 0.285$ and Spearman's $r_s = 0.091$, $p = 0.348$), were not statistically significant.

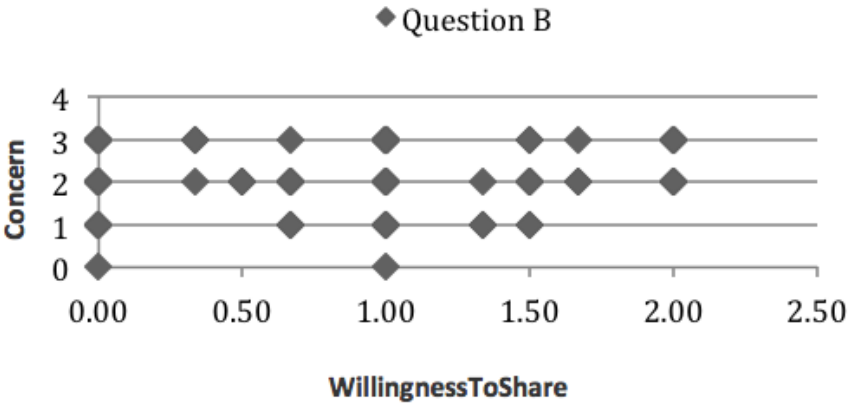


FIGURE 5.9: Plot between *WillingnessToShare* and *Concern* in Question B.

This finding supports the existence of the privacy paradox, since it shows that there is no strong correlation between people's attitudes towards privacy and their disclosure decisions.

5.2.5 Gender differences

As mentioned before, 90 participants were male whereas 60 were female. We wished to identify potential differences between men and women in their responses, and more specifically, whether women tend to be more private than men online, as several studies have shown (see Chapter 3). To that purpose, we investigated separately the answers of men and women in the scenario-based questions and revealed a few differences between them.

Wikipedia. Table 5.6 shows the responses of male and female participants to the Wikipedia scenario. It appears that a little over half of the male (58%) and a bit less than half of the female (43%) participants use Wikipedia. According to the table, most male participants were positive to the Wikipedia scenario (81%), however the majority of female participants was much more hesitant to share their location in this scenario (only 42% of female participants gave a positive answer). A chi-square test confirmed this statement ($chi - square = 15.214$ with $df = 3$, $p = 0.002$).

Wikipedia		
Answer	**Women**	**Men**
Yes	42%	81%
Maybe	12%	9.5%
No	35%	9.5%
Not answered	11%	0%
Total Responses	26	53

TABLE 5.6: Gender Differences in Wikipedia scenario.

Twitter. As Table 5.7 indicates the vast majority of female participants who replied to the Twitter scenario gave a negative answer. On the other hand, the responses of male participants varied greatly, with 39% answering "No", 25% "Maybe", and 34% "Yes". Approximately half of the male participants use Twitter (49%) whereas less than half of the female participants (37%) are Twitter users. A chi-square test showed that there is a relationship between the choice of answer in the Twitter scenario and the gender ($chi - square = 6.572$ with $df = 3$, $p = 0.087$).

Twitter		
Answer	**Women**	**Men**
Yes	23%	34%
Maybe	9%	25%
No	68%	39%
Not answered	0%	2%
Total Responses	22	44

TABLE 5.7: Gender Differences in Twitter scenario.

IMDb. Table 5.8 shows the gender differences in the IMDb scenario. The differences in the responses between the two genders were very small and according to the chi-square test there is no relationship between the choice of answer in the IMDb scenario and the gender ($chi - square = 0.677$ with $df = 3$, $p = 0.879$).

IMDb		
Answer	**Women**	**Men**
Yes	45%	56%
Maybe	9%	6%
No	36%	38%
Not answered	10%	0%
Total Responses	22	34

TABLE 5.8: Gender Differences in IMDb scenario.

Facebook. Finally, Table 5.9 shows the gender differences in the Facebook scenario. The majority of participants of both sexes gave negative answers and the chi-square test showed that there is no relationship between the choice of answer in the Facebook scenario and the gender ($chi - square = 1.440$ with $df = 3$, $p = 0.696$).

The results show that in comparison to male participants, female participants were much more reluctant to share their location in Wikipedia and Twitter. In the rest of the applications and their equivalent scenarios the gender did not have any relation with the choice of answer ("Yes", "Maybe", or "No").

Facebook		
Answer	**Women**	**Men**
Yes	13%	16%
Maybe	21%	30%
No	58%	52%
Not answered	8%	2%
Total Responses	53	73

TABLE 5.9: Gender Differences in Facebook scenario.

5.3 Qualitative analysis of survey results

Following the quantitative analysis of the survey results, we performed a qualitative analysis of the participants' justifications on the scenarios. For each scenario-based question, participants were asked to justify their decision in a single open answer. The analysis was based on the identification of themes followed by coding responses against those themes. In total we recorded 303 justifications across the five scenarios, with an average justification length of 18 words.

Quite often participants used the same or similar wording within their justifications. Figures 5.10 and 5.11 illustrate two examples of top word occurrences in the two applications that had the greatest number of responses. The first one shows the most popular word in the Facebook justifications, whereas the second one the most popular word in the Wikipedia justifications. These diagrams show that while some words are common, they are used in a wide range of ways, and although these diagrams provide interesting examples of people's reasoning, a more sophisticated thematic analysis was needed to locate patterns.

At first, a familiarisation stage took place where we established and developed a thematic framework for the analysis. For example, many participants stated that they would post their location on Wikipedia, simply because it is a helpful and convenient service, thus establishing the theme *Application benefits*. For many participants trust in the service was very important; for instance, many trusted Wikipedia but not Facebook — establishing a theme of *Trust in the application*. Each justification was then coded against these themes with refinements and/or extensions of the thematic framework where necessary. With the purpose of ensuring that their scope and content was clearly set out, all themes were given appropriate names and definitions. The final step of the analysis included a process of mapping and interpretation, clustering themes together in order to make sense of the responses in a holistic way.

FIGURE 5.10: Occurrences of the word "friends" in the justifications for the Facebook scenario.

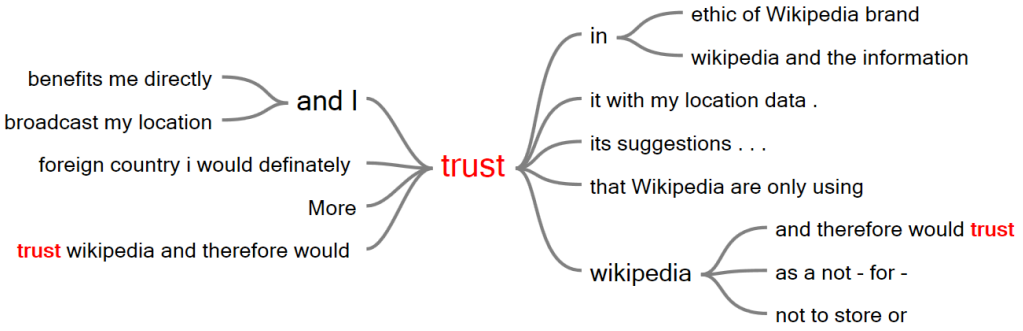


FIGURE 5.11: Occurrences of the word "trust" in the justifications for the Wikipedia scenario.

5.3.1 Interpreting the results of the qualitative analysis

During the final stage, themes were clustered together. To that purpose, we employed a two dimensional matrix. The horizontal dimension is a simple categorisation based on

whether the participant had answered positively, neutrally or negatively to the location-sharing question. The vertical was inspired by structuration, and the duality of agency and structure.

All the themes placed into the matrix are shown in Table 5.10. In the structuration dimension we placed each theme in one of three categories: Agency, Contextual and Situated Aspects.

Agency Aspects. The first category of *Agency Aspects* deals with people's opinions and views. Almost all the themes within this category reflected participants' negative opinions with regards to online location sharing. In other words, people who did not wish to publish their location tended to make justifications based on their general views, rather than some feature or aspect of the scenario. They were therefore acting largely as agents. For example, many people stated that they are simply not willing to publish their location online, some said that they are not interested in doing so, whereas others said that they do not feel comfortable with it:

"Because I don't like to tell the world where am I at any given time. "

Many also claimed that privacy and safety are the main reasons for their desire not to share online their location:

"I would not want to put information regarding my whereabouts on the internet. I feel this information could be misused could possibly result in a harmful situation for me. Also I like to keep aspects of my life private, I do not like to plaster my life across social networks for all to see."

Similarly, other studies have shown that people with high levels of privacy concern are less likely to engage in online sharing (Staddon et al., 2012). Several participants also stated that they saw no use from online location sharing, such as:

"I never use the location facility. I think using this is quite sad really! Why would you want to tell everyone where you are every minute of the day? You'd be checking in everywhere!"

Contextual Aspects. The second category *Contextual Aspects* contains themes that highlight how people's decisions were influenced either positively or negatively from the context in which they share their location with each application. They show the confluence of personal views with the specifics of the application, revealing how both agency and structure are used to reach decisions. Themes in this category mostly refer to contextual factors related to the application. For example, the theme *Existence of alternative options* refers to cases where the participants stated that they would not share their location with an application because there are better alternatives.

Matrix of identified themes			
Answer	**Agency Aspects**	**Contextual Aspects**	**Situated Aspects**
Yes	Comfortable with sharing (3)	Application Benefits (214) - Helpful application (51) - Convenient application (11) - Speed of access (6) Trust in the application (21) Location *not* visible to others (10) Location visible to others (20) No data manipulation/tracking by the application (9)	Sharing Experience with others (9) Benefits in specific scenario (5) Public event (4) Work-related event (3)
Maybe		Wishing control over data (10) Location used as metadata (in Twitter only) (2)	For fun/out of curiosity (10) Location visible to others (9) Weighing the benefits in scenario (7) Depends on type of event (6) - Private/public event (5) Depends on location (2)
No	Lack of interest in location sharing (14) Unwilling to share location (47) - Uncomfortable with sharing with others (5) Privacy Reasons (34) - Safety/security (8) - Sensitive locations (2) No benefits from location sharing (24)	Existence of alternative options (8) No trust in the application (16) Concerns over data manipulation (12) - Data useful only to companies (4) Location visible to others (8) Willing to share with friends only (4)	Private event (4)

Table 5.10: Themes developed through the qualitative analysis. The number of justifications coded for each theme is listed in brackets after the name of the theme.

Many participants, who gave positive responses in this category, stressed the benefits of the applications, and how helpful and convenient they are. The two main applications that received such justifications were Wikipedia and IMDb.

"Wikipedia makes no attempt to alter the pages I see based on a user profile. I get the same information augmented with some extra suggestions I can turn on or off that's fine providing it's my choice to use this feature."

Wikipedia was also considered a trustworthy application, whereas Facebook was considered not trusted.

"I don't trust Facebook, their security or their advertisers. I would rather keep this information to myself. My friends would probably already know where I am. Unfortunately my decision might be undermined by friends' activity."

Some participants also expressed concerns over their data being manipulated through Facebook and/or Twitter.

"It seems this data is more useful to companies using the Twitter API than to me or the people I'm communicating with."

In comparison to this, some participants were happy to use applications like Wikipedia, because, as they claimed, their data was not manipulated.

"Because I believe it's safe for Wikipedia not to keep my current location any further."

Significantly, for some people the fact that their location was not shown to other users in some of the applications (Wikipedia and IMDb) was a positive factor, whereas for other people the fact that their location was shared with others (in social network sites such as Facebook and Twitter) was a positive factor.

"It is more convenient than manually searching. This information is only revealed to Wikipedia rather than publicly."

"So that other people can see that I am at the event. It might save me some characters when writing the tweet because I wouldn't have to say where I was."

In the case of IMDb some participants stated that there are other alternatives to that application, they would prefer to use — e.g. Google search.

"Trust is the reason again. There are better ways to know about cinemas around without using an app like the IMDb. A simple search in internet will serve the purpose."

Situated Aspects. The third category, called *Situated Aspects*, refers to the aspects that influenced participants' decisions within the context of the specific scenario that was presented to them. In that sense, situated aspects include themes that deal with the specifics of the situation in the given scenario. In this category the majority of the

themes refer to participants whose answer in the scenarios was "Maybe". Participants would act under the influence of the set of structures given in the context of the specific scenario. All the themes deal with situated aspects of the scenarios, such as people's mood, or the location they are at, if it is a public or private, and so on.

"If the public event refers to a work-based event I would tweet my location. If it would be a personal activity no."

Many participants stated that they would publish their location, if they wanted to make it visible to others, for example their Twitter followers. Figure 5.12 shows the frequency of the word "event" in the Twitter scenario and highlights the effect that the nature of the event has on their decisions.

"I am happy to share my location when I am at a big public event which is probably attended by hundreds of others of people. I also only use Twitter for general posts which are suitable to be viewed by a wide audience. I never tweet private stuff. Therefore using my location in such tweets makes it easier for my followers to associated them with a place."

Some people wished to share their experience — as described in the scenario — with their friends in a social network:

"The place must be cool I'm excited about it and maybe I want to show the place to other friends."

FIGURE 5.12: Tag Tree with the word "event" in people's responses to the Twitter scenario.

Taking into account these three categories, it is evident that only a small number of people acted largely as independent agents, whereas the majority was influenced by the structures present in the scenario. This indicates that in practice most people tend to negotiate their privacy, and weigh the costs and benefits of the privacy trade-off depending on the context. In addition to this, the matrix clearly illustrates the dynamic nature of privacy decision-making. People decide to share their data online dynamically based on a set of contextual and situational aspects. This contradicts the assumptions behind the current online privacy settings which are static and do not take into account the dynamic nature of privacy disclosure mechanisms.

Answer	Agency Aspects	Contextual Aspects	Situated Aspects
Yes	3	274	21
Maybe	0	12	46
No	119	48	4

TABLE 5.11: Number of justifications coded in the themes.

Table 5.11 shows the total number of justifications that were coded in all the identified themes per cell. The table verifies the analysis; participants who acted largely as agents gave negative responses, whereas the responses of the participants, who were influenced by the structures, were more dispersed among the three possible answers and tended to be more positive in their answers.

5.4 Conclusion

In this chapter we presented a survey designed to investigate whether the privacy paradox holds for location data, and to explore the reasoning behind location privacy decisions. With that in mind, we performed a quantitative as well as a qualitative analysis of the survey data. The results of the quantitative analysis suggest that there is no correlation between people's stated views on privacy and their privacy decisions within a number of every-day scenarios. This is evidence that the privacy paradox *does* apply to location data.

We also found gender differences regarding privacy decision-making, as female participants were less inclined to share their location in applications such as Wikipedia and Twitter. These findings can be justified by current research on gender and online sharing — presented in Chapter 2 — as there are several studies that support that women are more privacy concerned than men (e.g. Wills and Zeljkovic, 2011; Hoy and Milne, 2010; Fogel and Nehmad, 2009; Youn and Hall, 2008).

We also performed a thematic analysis of the participants' justifications of their privacy decisions, in order to illuminate the decision making process. The analysis shows that privacy decisions can be seen as part of a process of structuration, where attitudes and values (people's free agency) are tempered by situation and context (external structures). The coding also suggests that agency is often a negative influence on sharing, whereas structures tend to be a more positive influence. In turn, this could potentially justify the discrepancy between people's stated attitudes and their actual disclosure behaviour.

In discussing structuration Giddens refers to the *duality of structure* meaning that structure can be both the medium as well as the outcome of activities that recursively take place. Seeing privacy decision-making as a process of structuration also implies that agency leads to new structures, in other words that decisions establish new norms that become new influencing structures. In that sense, privacy decision-making can be seen as a cycle, where people expect certain rules to apply when they share information about themselves and therefore new *rules* and *resources* may be developed. A potential example could be the introduction of Facebook's audience selector tool which offers users the ability to manage their audiences every time they make a sharing decision since there was an obvious need for better managing audiences per post.

The analysis presented in this chapter, and the role it implies for structuration, informs the on-going work of Privacy by Design (Cavoukian, 2012) and has implications for the design of privacy systems. More specifically, our analysis questions the way that privacy preferences are currently recorded, since it implies that privacy preferences should not be static, as users' sharing decisions are dynamic and dependent on external structures that only become apparent in a given context. In the following chapter we aim to extend this analysis through a series of focus groups to verify the survey outcomes, explore in more depth the role of context in privacy decisions, and also highlight potential gender differences.

Chapter 6

An In-Depth Study into Disclosure Decisions

In the previous chapter we presented the outcomes of an online survey that aimed to gain an understanding of people's online sharing decisions, with a particular focus on the theories of the privacy paradox, the privacy trade-off and structuration. The study provided evidence for the existence of a privacy paradox regarding location data and highlighted that privacy decisions are heavily influenced by contextual factors.

In this chapter we aim to achieve a deeper understanding of people's disclosure decisions through a series of focus groups. We also seek to refine and clarify the role of context within privacy decisions, as well as investigate in more depth the findings of the previous study. This study consists of several steps: *a*) designing the focus groups, *b*) organising and running separate sessions, *c*) the data transcription, and *d*) the analysis of the data. All steps of the process, as well as their outcomes are described in the following sections.

6.1 Using focus groups as a follow-up study

This study comprises of a series of focus groups that act as a follow-up to the survey and aim to explore how people articulate their location sharing attitudes. More specifically, our aim is to explore in more detail the contextual factors identified in the survey analysis by gathering qualitative data during the sessions. The reasoning behind this combination is based on Morgan (1996) and his framework on combining focus groups and surveys as complementary research methods. According to Morgan, focus groups can offer deeper insight into how the participants discuss the topics of a survey (since a survey has a limited set of questions). Apart from this, the purpose of both studies (online survey and focus groups) is to contribute to the development of a conceptual

model of privacy. In that sense, the practical outcome of the current study is to provide more input for the design of the model.

Focus groups constitute an established research method that can be used to elicit detailed information about a certain subject. A survey is able to provide a large number of responses, yet these responses are usually close-ended with limited options for the respondent to select from. On the other hand, focus groups offer a means to explore a topic in detail. With regards to privacy decision-making, they can reveal critical differences among the participants concerning the conditions under which a privacy decision was made. The most important advantage of focus group studies is the fact that they are based on the interactions between the participants, who exchange opinions and experiences and comment on them. As focus groups explore people's knowledge and experiences, they also offer a means to explore not only people's thoughts but also how they reason their thoughts and why (Kitzinger, 1995). This can be extremely useful when studying privacy attitudes, as they have the potential to offer interesting details regarding people's reasoning of their privacy attitudes and disclosure behaviours.

We conducted three separate focus group sessions, each one with a different set of participants (in total 19 participants). The participants in the first two sessions were undergraduate students of the University of Southampton, aged between 18-21 years old; from fields unrelated to Computer Science and who use location-aware smart enabled devices. The first session included only female participants, whereas the second only male participants. The third focus group session took place with participants who are postgraduate research students of the Web Science Doctoral Training Centre. All participants of this group conduct interdisciplinary studies related to the Web (topics around privacy, open data, cybercrime, and security), and have good knowledge of the ongoing debate around privacy.

There are two main reasons behind this recruiting approach. Following the findings of the online survey, where we found that women were less inclined to release their location in several Web applications (such as Twitter and Wikipedia), it became apparent that fruitful findings may come up from conducting separate sessions for women and men. In addition to this, we hoped to explore potential differences in privacy attitudes and behaviour between people who think actively about privacy (in this case in their research and study) and young people who are more or less enthusiastic users of web-based applications.

In the first two sessions (with undergraduate students) participants were also requested to fulfil the following criteria:

- use the Mobile Web in their daily life (i.e. users of smartphones, and tablets)
- publish their location online, and

- are frequent users of social network sites (e.g. Facebook, Twitter).

The first focus group consisted of 7 female participants, the second of 6 male participants, whereas the third one consisted of 6 research students. The duration of each session was approximately one hour.

Each session focused on a set of questions about participant's online sharing decisions, hoping to uncover the reasoning behind their decisions, and to explore any differences in privacy decision-making from different demographics. With the aim to observe how participants would discuss and interact on topics covered in the survey, we employed a similar approach. Hence, the largest part of each session included different real-life scenarios aimed to explore the privacy trade-off in practice — similarly to the scenarios of the survey. Participants were prompted to answer the question *"Would you share your location in this scenario?"* with "Yes", "Maybe", or "No" and afterwards justify their responses within the context of a discussion. The study gained ethics approval by the University of Southampton Ethics Committee [1] (Ethics reference number: 5482).

In more detail, each session was structured in the following format:

- An introductory discussion regarding the subject, where the moderator used a PowerPoint presentation to make the participants familiar with the topic of discussion and then asked the participants to fill in a short questionnaire (both of them are presented in Appendix C). The purpose of the questionnaire was to profile each participant with regards to their location sharing attitudes and their use of privacy settings.
- Three scenarios — one on Facebook, one on Twitter and one on Wikipedia, similar to the ones included in the online survey — were presented to the participants through the Powerpoint presentation, asking them whether they would share their location in the scenario presented. A discussion followed during which participants were asked to justify their responses. Apart from this, the discussion focused on the importance of context in the participants' willingness to share their location online.
- Wrap-up of the session.

Following the focus group sessions, all data was anonymised and transcribed into digital format. A thematic analysis took place using NVivo 9 based on the six-step approach proposed by Braun and Clarke (2006). After getting accustomed with the data, the coding process began during which we identified patterns within the data. Once the list of coded data was produced, we began identifying themes. Then, the themes were evaluated, and if necessary we went back to the original data and developed new themes

[1] www.ergo.soton.ac.uk

(and sub-themes). The next step involved the appropriate naming and definition of the themes, to ensure that their scope and content were clearly set out. Finally, the emerging themes from the coding process were grouped together and formed a set of categories. The categories were analysed in terms of their relationship with the topic and contribution to the understanding of the data.

6.2 Initial Findings

The first part of the session involved a questionnaire that was handed to the participants to fill in, containing the following Likert-scale questions:

- **Question A.** Are you satisfied from the privacy settings offered to you by web applications?
- **Question B.** Most of the broadly used privacy settings are based on the concept of "who has access to see your data". Are you satisfied with them?

Participants were prompted to choose between a range of five different answers: "Agree completely" - 1, "Agree" - 2, "Neutral" - 3, "Disagree" - 4, and "Disagree completely" - 5. Table 6.1 shows the mean answers to these questions per group.

Group	Question A	Question B
Male student	2.0 (Agree)	2.0 (Agree)
Female student	2.1 (Agree)	2.3 (Agree)
Research student	3.8 (Disagree)	3.2 (Neutral)

TABLE 6.1: Mean answers to questions A and B per group.

Discussing whether they are satisfied with the current privacy settings, the research students appeared to be rather dissatisfied expressing that they do not understand them:

"I'm unhappy more because I don't understand, rather than because they may or may not satisfy my needs."

For the final two Likert-scale questions the different options were: "Never" - 1, "Rarely" - 2, "Once a month or so" - 3, "More than once a week" - 4, and "All the time" - 5. The mean answers to these questions are displayed in Table 6.2.

- **Question C.** In your mobile device do you ever turn on the location settings?
- **Question D.** How often do you post your location in a social network (e.g. Facebook, Twitter, Google+)?

Group	Question C	Question D
Male student	4.7 (All the time)	3.2 (Once a month or so)
Female student	3.9 (More than once a week)	3.4 (Once a month or so)
Research student	3.4 (Once a month or so)	2.0 (Rarely)

TABLE 6.2: Mean answers to questions C and D per group.

Although quantitative outcomes cannot be produced through focus groups, the purpose of the questionnaire was to conduct an initial screening of the participants in each session. It appears that the participants of the first two groups (male and female undergraduate students) were more satisfied with the current privacy settings and at the same time they were more eager to share their location online than the research students. These findings are somewhat expected, as the research students actively study topics related to privacy and security, hence it is sensible to assume that they might question more the efficiency of current online privacy systems than any other user of these systems.

As described briefly in the previous section, the main part of each focus group session consisted of a discussion around the participants' location sharing decisions in different scenarios. The following subsections describe the findings of the qualitative analysis of the focus group data.

6.2.1 Scenario-based questions

During each session three scenarios were presented to the participants:

- a Facebook scenario where they were asked whether they would share their location if they were at the airport about to go on holidays,
- a Wikipedia scenario asking them whether they would share their location with Wikipedia in order to explore interesting things around the campus of the university, and
- a Twitter scenario where they were asked whether they would share their location with their Twitter followers if they were at a concert.

Each scenario was presented through a small description along with some visual aid (e.g. pictures from the user interface of each application, which are included in the handout that was given to the participants, see Appendix B). Participants were expected to answer "Yes", "Maybe" or "No" and then justify their answer in the context of a discussion.

Answer	Facebook	Twitter	Wikipedia
Yes	11	4	15
Maybe	5	7	4
No	3	4	0
No Answer	0	4	0

TABLE 6.3: Answers to scenarios from all the participants of all three groups.

Table 6.3 illustrates the answers of the participants to these questions. It is evident that in the Facebook and Wikipedia scenarios people were willing to negotiate their privacy and potentially share their location online. In the case of Twitter several participants either stated that they do not know how to post tweets with their location or that they would write where they are (or what they are doing) in the 140 characters of their tweet (instead of using the geo-tagged functionality of Twitter).

Note: In the Twitter scenario there were a few participants who did not have a Twitter account, therefore did not answer the question.

6.2.1.1 Analysis per application

Facebook Scenario. *"You are travelling abroad with a friend of yours (also your Facebook friend) and at the moment you are at the airport. Would you post your location on your Facebook wall?"*

Facebook			
Group	**Yes**	**Maybe**	**No**
Male student	4	2	0
Female student	6	1	0
Research student	1	2	3

TABLE 6.4: Participant answers to the Facebook scenario.

The answers from the participants of each group are presented in Table 6.4. As the Table illustrates, the majority of the male and female undergraduate students were willing to share their location on Facebook in the scenario given.

Twitter Scenario. *"You are attending a concert in London and thinking about tweeting about it. Would you tweet with your location?"*

Twitter			
Group	**Yes**	**Maybe**	**No**
Male student	2	2	1
Female student	2	0	3
Research student	0	5	0

TABLE 6.5: Participant answers to the Twitter scenario.

In this scenario most participants — male and female — argued that most probably they would not tweet with their GPS location on, but they would rather state within their tweet where they are. As stated earlier, some female participants argued that they did not know how to use the geo-tagged functionality of Twitter.

Interestingly, all research students (who are Twitter users) would debate tweeting with their location depending on the scenario.

Wikipedia Scenario. *"You are visiting Southampton for the first time with a friend and wish to visit the sights of the area near the university. Your friend suggests using Wikipedia, as it shows on a map links to the Wiki pages of all the nearby sights. Would you allow Wikipedia to determine your location?"*

Wikipedia			
Group	**Yes**	**Maybe**	**No**
Male student	5	1	0
Female student	6	1	0
Research student	4	2	0

TABLE 6.6: Participant answers to the Wikipedia scenario.

This scenario was by far the one where participants felt more comfortable to share their location. Therefore, most of the participants from all three groups answered "Yes".

6.3 Qualitative analysis

Following their answers to each scenario-based question, the participants of all groups were asked to justify their answer in front of the rest of the participants and the moderator. In this section we will go through the outcomes of the qualitative analysis of these discussions. All themes that were developed through this analysis are presented in

Table 6.7. Following the coding stage, the themes were grouped together to form three separate categories: social capital, trust, and functionality.

Identified Themes	
Category	**Themes**
Social Capital	Self-presentation (24), Interacting with others (7), Who sees my location (9), Interesting location (out of the ordinary) (18), Apathy (not interested in location sharing) (5)
Trust	Debating trust /Not sure (28), Trust in the app (15), Not considering trust (3), Rational Ignorance (11), Valence effect (10), Lack of control (11), Concerns over data (4), Priv. settings issues (16), Suggestions for priv. settings (5)
Functionality	Application Benefits (21), Benefits in this scenario (17), Costs in scenario (4), Safety issues (8), Implicit Location in Twitter (not geo-tagged) (8)

TABLE 6.7: Themes developed during the analysis.

6.3.1 Social capital

For the undergraduate participants their social circle has a strong influence on their sharing decisions. Participants use social network sites as a means of interacting with their friends, thus the equivalent theme was developed (*Interacting with others*). For example, a female undergraduate student suggested that:

"It's nice to interact with people, let's say if you are in a group and they are there as well it's nice to have a look back."

All of the undergraduate participants enthusiastically shared their location with their friends especially when they were at a "cool location" — for example, they stressed that they usually share their location during a holiday abroad, but not when out having a normal lunch. Some research students also pointed that they would share their location at an out-of the ordinary location (e.g. conference). Based on these reports, the theme *Interesting Location* was created. For instance, a male undergraduate participant justified his location sharing by saying that:

"So that people know I'm going to an exciting place."

In addition to this, a female student explained in more detail her reasoning behind this:

"Well, it's fun like when you post something you want to get lots of 'Likes' and stuff like that. If you gone to a cool place, you want people to be like.. I don't know, if you are about to go travelling you'd hope that some people to respond, by liking or say 'Oh

I'm so jealous' or something. You know, I wouldn't post a status about, if I was like in Costa or something, cause no one is going to get excited about that. So there will be no point. But, if you are going away somewhere, travelling or something, you'd hope that lots of people could see it and responded in a positive way... I guess."

Our findings are in agreement with a study consisting of interviews and a survey targeting people who use location-based applications, which found that people wish to share their location in places outside of their daily routine (Lindqvist et al., 2011).

Students mostly ignored the *invisible audiences* (boyd, 2008), their rationale focused on their friends that would see their data, leading to the development of the theme *Who sees my location*. To them, disclosure decisions seemed to mostly depend on who would have access to their data in terms of specific people — e.g. friends and family, however some participants also mentioned job recruiters. For instance, some undergraduate participants stated that sometimes when they make disclosure decisions they think about specific people they would rather keep their location secret from — e.g. their mother or a specific friend. For example, a female student thought of her mother and stated that:

"If my mum doesn't know I'm going somewhere, I don't want to be caught out."

Several studies have explored the dynamics between different types of relationships and their presence in social network sites. Through a series of focus groups, Fusco et al. (2012) addressed the question with whom people would willingly share their real-time location in a social network site. Five separate types of relationships were investigated: family members, friends, people from working environment, commercial, and government relationships. The study had a special focus on interpersonal trust and found that location-based applications can potentially have a negative impact on trust in different relationships.

Several researchers have specifically studied location sharing and its effect on family dynamics. Mancini et al. (2011) used a location-tracking application to study this relationship in a qualitative study with two different families. They found that tensions between family members are highly likely to take place due to the affordances of location sharing. Another study found that although such technologies may be used as a means of *digital nurturing* (i.e. ensuring safety of other family members), they have strong potential to result in a *domestic panopticon*, where family members spy upon each other and therefore trust between them is undermined (Boesen et al., 2010). Bearing that in mind, it becomes clear why being able to control *who* has access to one's data online is important to individuals.

Apart from being watched by family members, participants discussed the possibility that people, with whom they have a professional relationship, might have access to their location data. For instance, a male student discussed the possibility of academic staff viewing their location:

"You don't want to, let's say lecturers, if you know that one of your Facebook friends is a lecturer then you wouldn't want to share some of the location."

This issue arises due to *context collapse*, a characteristic of many online applications, such as Facebook, where many different audiences are merged into one single group (boyd, 2008; Vitak, 2012).

Some of the participants also showed a few concerns with regards to their online presence when it comes to job seeking. They were aware that some companies check people's online presence during the screening process, one male student stated that:

"It's the whole thing about certain businesses check your Facebook during a job application. Since then I've actually untagged myself from certain photos drunk.."

The research students also discussed the matter of companies having access to potential employees' online presence:

"Can't you, when you start applying for jobs just block stuff from people who aren't your friends? Because by the time you start looking for a job you realise 'Oh yes, maybe all that stuff I posted when I was younger'."

Similarly to the participants of the focus groups, Hargittai and Litt (2013) found that young people are likely to change their privacy settings with the expectation that potential employers might look up for more information about them. Their study also revealed that women and people with stronger online privacy skills (i.e. the level of understanding of several terms related to online privacy management) are more likely to do so.

Participants from all three groups wished to share their data as a means of *self-presentation* and consequently the equivalent theme emerged. It was striking how willing all young undergraduates were to publish their data in social network sites, especially when they were involved in an activity that would attract people's attention. Some undergraduate students confessed that they do it for "showing-off":

"If you are about to go travelling you'd hope that some people to respond by liking or say 'Oh I'm so jealous' or something."
and *"I suppose it's just showing off."*

Previous research has also examined the relationship between young people's narcissistic traits and self-disclosures online and found that higher levels of narcissism are associated with more frequent self-promoting disclosures (Buffardi and Campbell, 2008; Mehdizadeh, 2010; Ong et al., 2011). The research students stated that they often do it for professional reasons, yet they admitted that they care about how others view them (one research student referred to this as "brand management"). In the literature this is called *impression management* (Goffman, 1959; boyd, 2008). As already mentioned in

Chapter 3, self-presentation is a concept that has long been studied. Erving Goffman introduced it by affirming that people wish to control the way others view them during their interactions (Goffman, 1959). The Social Web has offered an ideal ground for people to present themselves to others, as it happens with great ease — with a click of the mouse — and it may be projected to multiple audiences simultaneously. According to the focus-group participants, self-presentation was a strong positive factor for location sharing.

A few participants stressed that usually they are not interested in sharing their location in social network sites, arguing that nobody really cares about where they are. A male undergraduate participant claimed that this is due to lack of interest:

"Because I don't care for anyone to particularly see it. I don't think people want to see where I am at or whatever, so no I'm not bothered about it. I'm not interested."

Similarly, a research student suggested that this happens due to apathy. As a result, the theme *Apathy* was developed:

"A lot of my decisions to post my location are down to apathy... it's not a privacy concern it's more of that I don't think anyone really cares or I can't be bothered to put the location on."

As described earlier in this section, for many participants publishing their location online to attract the attention of their friends often depended on the location they were at. They would share their location when they were at an interesting place — e.g. on holiday, or at a concert — out of their ordinary life. Research students would share their location at a public event often for professional reasons. It appears that *context* played a key role in their decision-making process, *ergo* depending on the context they would decide to share or not.

All the themes presented in this section were grouped together under a single category called Social Capital. As discussed in Chapter 2, social capital is a concept that deals with the benefits that arise from people's social interactions (e.g. emotional benefits, new information, and so on). The diagram presented in Figure 6.1 shows all the themes that are part of this category.

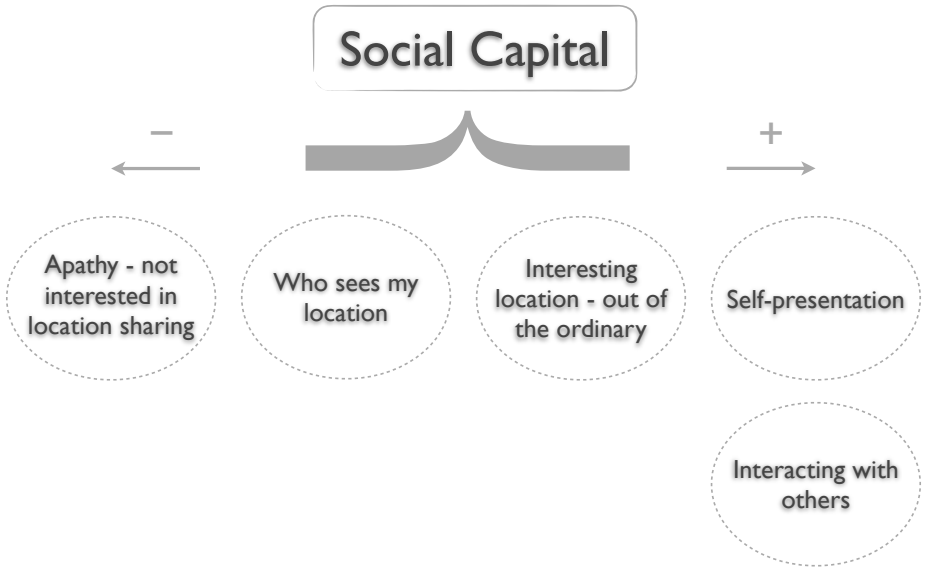


FIGURE 6.1: Themes under the 'Social Capital' category.

6.3.2 Trust

The concept of trust in the applications that handle people's data was evident in all the focus groups. However, the participants of the different groups approached this matter differently. For instance, the undergraduate students had an entirely different viewpoint on the matter to the research students. During the sessions with the undergraduate students trust was a challenging topic, as participants were not willing to provide a lot of input. The most probable reason behind this is that, as participants admitted, they do not actively think on trust in the application when they make privacy decisions. A male student stated that he trusted online applications suggesting that:

"Yeah, I'd probably say so. They've done nothing for me to not trust them."

A female student suggested that there have not been any privacy violation cases online, a statement that indicates possible lack of knowledge or understanding around this topic:

"You think that, like you know, if there was anything dodgy going on it would have been hopefully found out by now."

A few female participants stated that trust to the application was a topic that they had not considered before, thus the theme *No consideration about trust* was formed. The rest of the participants stated that although they were aware of the general privacy concerns relating to these applications they did not pay so much attention to them, as they do not understand many of their aspects, so they decided to ignore thinking about the issue of trust altogether. Several male participants stated that the benefits clearly outweigh the costs. Undergraduate participants actively decided to ignore the issue, rather than

learning more about it, a concept called *rational ignorance*. For instance, a couple of female students suggested that:

"People don't really understand, like the security of it as much, so it goes over your head the idea yeah." and
"I think it's more like Ignorance is Bliss. We don't really care. [...] I do hear news stories but I can sort of ignore them."

Apart from that, participants stated that nothing bad has ever happened to them, a concept known as the *valence effect*. Example quotes on this topic from female participants are the following:

"Most of the stuff I post about is not that interesting anyway, so.."
"It's unlikely that something bad will happen."

A common claim from their part was that the things they share online are of minor importance to anyone. Other researchers have suggested that both biases may influence privacy decisions (Acquisti and Grossklags, 2007).

As stated before, the most important thought that would cross the undergraduate participants' minds when sharing their data online was *who* would see their data. Participants did understand that on Facebook they shared their data with their friends (or anyone, as they did express doubts about the visibility of their data), on Twitter with anyone (unless they have a private account), but several participants from all three groups showed lack of awareness regarding Wikipedia. As a result, a few of them wondered whether by allowing Wikipedia to access their location, their data would be broadcast to their online friends. That can be explained by the fact that they were all regular sharing users of the first two applications but did not generate input to Wikipedia. For instance, a research student stated that:

"Whereas when I think about Wikipedia I don't know anything about what they do with my data. In a way you can have trust in an application, even if trusting it in a way that you know it's going to do something horrible with what you are putting in there, at least you know what's happening. With Wikipedia I don't think I would give them my location, cause I just don't know."

Privacy management was another issue discussed during the focus group sessions with both female and male undergraduate participants. The opinions varied; many did not show a lot of interest in the topic, however some of them had experiences where information about them (location, photos, or activities) was shown to online friends without their knowledge.

For example, in a discussion between the female students, two of them stated that they do not often change their privacy settings:

A: *"Mine was set to private, but I never set it to private again, so it's probably non-private (laughs)."*
B: *"Yeah, everybody's probably like that!"*

A few of them expressed ideas about changing current privacy settings to make them simpler, hence the theme *Suggestions for priv. settings* was created. A female student suggested simplifying the privacy settings options:

"I think it would be easier if it was just public or private, but then how it is probably suits more people cause it's more flexible."

Another female student referred to the issue of the frequent changes in the privacy settings of Facebook: *"I think that they should notify you if your privacy settings have changed..".*

The research students also discussed the issue of transparency:

"I would just want perfect transparency over exactly what I'm sharing with who up-to-date. So, it's not so much what I'm sharing, sharing information, I'm happy to share but I just want to understand it and I don't understand it and I really do think that you can see other people's profiles but I can't see myself objectively what other people see about me. So, like an app where I can do that but I want perfect transparency and easy terms to understand what actually means."

As this quote indicates, in contrast to the undergraduate participants, the research students had strong opinions regarding online trust. They were all up-to-date with the latest advances relating to privacy and security issues and were rather critical about them. Participants were aware of the privacy issues regarding Facebook and appeared to be more relaxed regarding trust in Twitter and Wikipedia. The fact that Twitter is completely public made the participants more assured about what to expect from it as an application. Wikipedia was by far the most trusted application of the three. The research students did not show any signs of rational ignorance or the valence effect. Instead some of them acknowledged as a matter of concern the fact that they feel unaware of the consequences of their privacy decisions, especially the long-term consequences:

"So, I think most of us go around thinking we have a partial way of understanding the ways in which the context initiates publishing information about ourselves."

Nonetheless, participants from all three groups shared the opinion that they need to feel more accountable of the affordances of their data. They all discussed, for example, the constant changes in the privacy settings of Facebook and showed discomfort about that. The research students expressed their opinions on this matter more strongly (as discussed above), yet the other two groups expressed certain annoyance as well, particularly when they do not feel in control of their online data. For instance, in a discussion about Facebook privacy a couple of male students stated that:

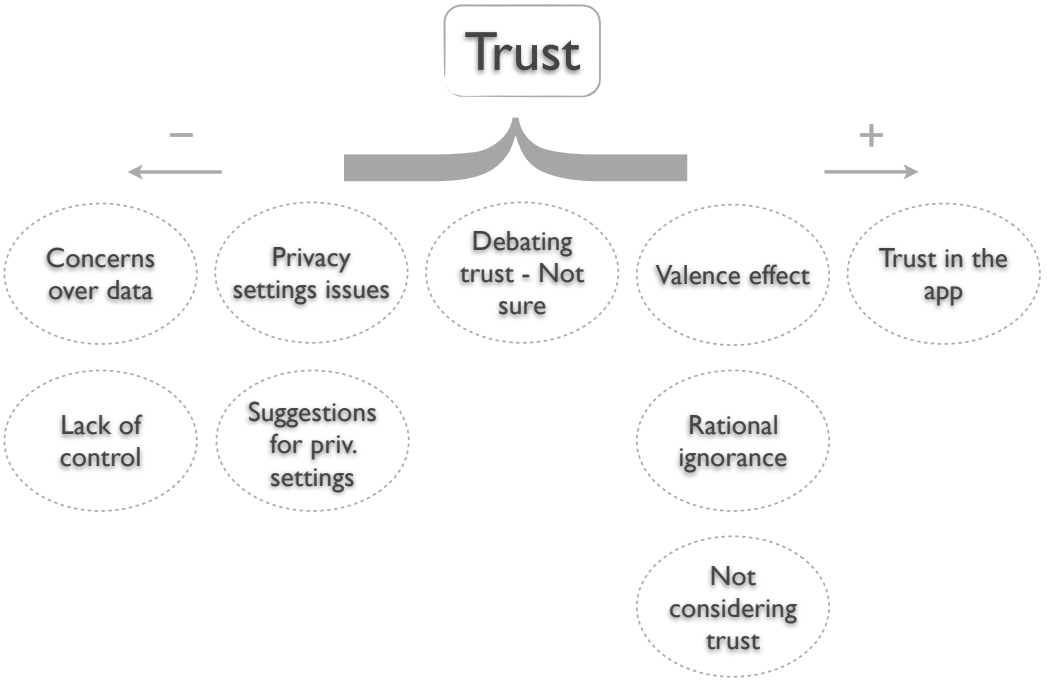


FIGURE 6.2: Themes under the 'Trust' category.

"And also they start to... in the Timeline [2]*, in the history, they are starting to implement that... where you can actually check in 2010 where you've been, so maybe that's not good when someone else is checking."*

"They're not that clear sometimes about differences between like posts and other things. Sometimes I've seen, I've just done something set public and I don't really like things being public. I've had to change it manually myself. I'd rather they made it a little bit more simple and actually privacy settings homepage about how we can specify.."

All the findings presented here were grouped together under a single category called Trust (see Figure 6.2).

6.3.3 Functionality

It is evident from the responses of the research students that they usually base their decisions on the benefits they would receive in return. Their justifications were usually based on the pay-off the trade-off would have for them in the specific scenario. As a consequence, they share their data when the perceived benefits outweigh the costs:

"Less risk with being away from home too long, and anyone who chooses to follow my twitter account would be home, interested."

[2]Facebook feature, part of the user profile page

Still, some of them did admit that at times their decision is rushed and task-focused:

"But I think it's decision making, rushed decision making and we have different hats on different roles that you take on as different people. Sometimes I think it doesn't really matter what my friends say."

They were much more positive in their answers in the Wikipedia scenario, stating that allowing Wikipedia to access their location would have clear benefits to them. Of course their decision was based on the benefits in tandem with the fact that they trust Wikipedia more than other applications:

"I'm tit for tat when it comes to things like this. So, the Wikipedia example is a good example of where I would give out my location because I'm getting something back instantly. Whereas sharing my location on Facebook what do I get from that? I don't get anything, so why do I do it?"

The undergraduate students did not appear to have much experience of trading data for practical gains, except for specific cases where the benefits of the trade-off were rather explicit. Wikipedia was a good example, as is not a social network and has clear advantages to them (the context of the scenario assisted to this):

"In this situation you wouldn't need an account, wherever would be your location, it's nothing, but when I say nothing, they can track you not you as a person but your location at the time yes, but it's not related to me, it could be anyone."

Participants of the female undergraduate group also stressed the advantages of sharing their location in exchange for vouchers or directions:

"You are to redeem vouchers, you have to say where you are, like if you are with O2 [3] *and you want to get a deal or something you have to show your location to see where the nearest deals are."*

A study on location sharing practices conducted by Patil et al. (2012) also confirmed that people share their location in order to receive rewards for "checking-in".

Finally, a few research students but also a couple of male participants stressed *safety issues* when posting their location, thus the equivalent theme was created:

"In that scenario, I would agree with participant X that I wouldn't post I was in the airport because it would imply that my house is empty. So, that would put me off doing that."

All the above-mentioned findings were coded together under a common category called Functionality and are displayed in Figure 6.3. Functionality is defined as the quality of having a practical use, in our study it refers to cases where there is a practical gain

[3] Telecommunications, internet and financial services provider in the UK

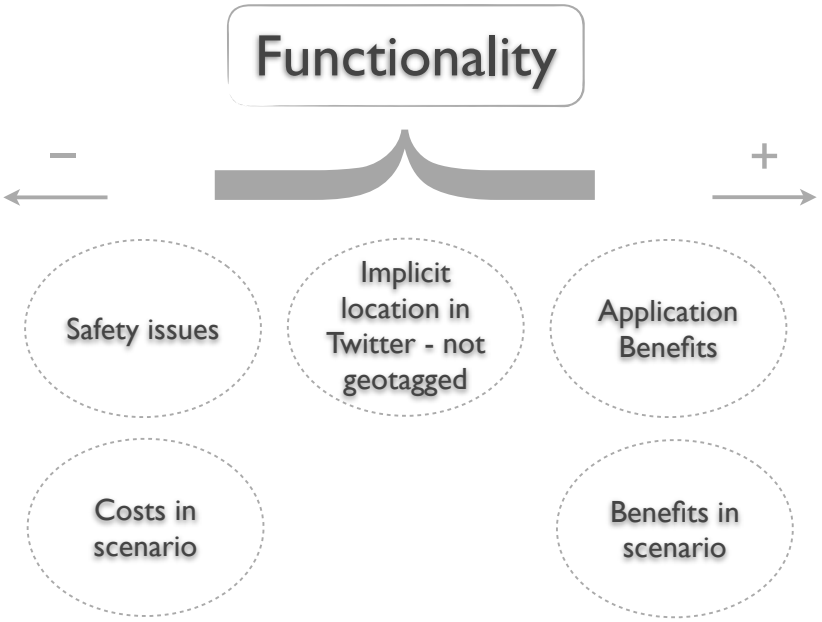


FIGURE 6.3: Themes under the 'Functionality' category.

from trading one's information, usually in the form of a functionality of the application (such as receiving info that is filtered by location) but occasionally in monetary terms (e.g. by sharing their location on twitter they would win vouchers). It is worth pointing out that social capital was developed separately, rather than as a functionality, this is because of its importance in terms of the participants' thinking, but also because it does not always provide a concrete practical outcome from its trade-off.

6.4 Revisiting the theory of cognitive dissonance

The justifications of the undergraduate students can also be associated with the theory of *cognitive dissonance*, presented in Chapter 2. This theory refers to a state of tension that people experience when their decisions are in conflict with their belief system. When faced with such situations people naturally attempt to find a balance between their decisions and their beliefs in order to minimise the dissonance, either by changing their beliefs, or their decisions, or by changing their own perception of their decisions. The justifications of the participants can easily be seen as attempts to reduce the dissonance caused by their positive decisions to location sharing and their privacy-related concerns; in fewer words, the tension caused by the privacy paradox.

For example, some students stated that the overall benefits significantly outweigh the costs:

"I think always the benefits of these things completely outweigh the tiny risk that you are ever exposed to, but then that might be me being naive but I wouldn't associate any risk at all really."

As we already pointed out, some students undermined the importance of their shared information:

"Most of the stuff I share is not that interesting anyway."
and
"Some people can guess your location anyway."

The use of such justifications can be seen as an attempt to rationalise their behaviour, and in turn reduce the dissonance. In the case of our participants, through these justifications online location sharing is consistent with their ideas around privacy and trust in online applications.

Still, someone may argue that these justifications are simply indications of apathy towards privacy. However, the participants indicated that they care about their privacy, especially their social privacy. As mentioned before, most of the participants cared about *who* would access their data and in some cases they did not wish specific people to access them. The question, however, remains when it comes to privacy and trust in online applications. As we will describe in more detail in the forthcoming section, the research students appeared to be more confident in their discussion on trust in web applications, whereas the undergraduate students appeared not as confident. In addition to this when it came to trust, the undergraduate students showed signs of rational ignorance and the valence effect. This implies that their justifications were not based on apathy but on a set of behavioural biases. Still, the relationship between cognitive dissonance and the privacy paradox needs to be studied in more detail.

6.5 Comparing the three groups

Apart from the above-mentioned observations, a number of commonalities and differences were identified among the three groups. We observed no strong differences in people's attitudes and behaviours between participants from different genders, however there were strong differences between undergraduate students and research students.

One notable difference is that the research students articulated their opinions in a sophisticated way using privacy-related vocabulary, while the undergraduates struggled to describe their reasoning or express their views. The research students highlighted a number of known privacy issues (e.g. social issues such as teenagers' sharing behaviour in social networks and related risks, legal issues, as well as news stories) revealing that they were well informed around numerous topics around privacy and security. On the other hand, the discussions between the participants of the other two groups indicated

that it was not easy for them to articulate their thoughts on privacy. Many of them showed lack of confidence in discussing the topic. It became clear that they lacked the vocabulary to discuss their perspectives fully, as they often began to talk but did not complete their sentences. From this, one might expect the research students to be rather hesitant to share their data online and consequently their answers in the three scenarios to be negative. However, when presented with the scenarios, they all tended to negotiate their privacy depending on the benefits they would receive. Some of them acknowledged that they think differently about privacy when they act as professionals than when they actually make privacy decisions themselves in a specific context. In the second case they admitted that they think as users and decide to trade their privacy over convenience.

In complete contrast to the undergraduate students, the research students took into account the role of third parties (i.e. applications that may access their data), whereas to the undergraduate students third parties were invisible. In that sense, the research students felt that the degree to which people are in *control* of their data is debatable.

"Cause that only controls which of your friends can see that information. Facebook still has it and can still use it when you've been tagged, they just won't show your mother."

Studying a privacy related subject appears to have a strong impact on people's privacy attitudes and it enables them to make a more conscious privacy trade-off. Yet, all the participants from all the groups admitted that not only do they share their data in different contexts, but they also share their data in social networks as a means of maintaining their social capital.

6.6 Discussion

In this chapter we undertook a detailed analysis of a series of focus groups, which showed that rational ignorance and the valence effect affect people's decision making about sharing their location.

According to the findings of our study, there were no strong differences in people's attitudes and behaviours between participants from different genders, however differences between undergraduate students and research students were evident. Undergraduate students cared more about who would see their data rather than what applications did with their data. Similarly to our findings, Young and Quan-Haase (2013) found that young students do manage their social privacy (how others view them), yet they are not concerned about how applications handle their data. Johnson et al. (2012) showed that 37% of their survey respondents did not wish specific people to see their profiles or content posted about them. Page et al. (2012) explained that this happens because people wish to control the boundaries of their relationship with these people (e.g. actual relationship with co-workers).

On the other hand, the research students focused on the greater picture regarding what happens with their data. They commit themselves in a trade-off, during which they think about the consequences of their privacy decisions — although they did admit that their decisions might often be rushed. Regardless of their different approaches towards sharing, participants from all groups based their sharing decisions on a variety of contextual factors.

A factor that influences people's disclosure decisions significantly and appeared in the justifications of the participants of all three focus groups is *context*. During the analysis it became evident that all the themes were related to context. For instance, participants would share their location if they were at an interesting location (out of their every day life) but not at the local coffee shop. In case any contextual factors changed (from application-related to scenario-specific factors), their privacy decision could easily change as well. In that sense, context played a primary role in participant's decisions, as in different contexts participants could have different disclosure behaviour. As a consequence, there were no themes developed specifically about context, yet context was a key factor in all themes.

Sharing decisions were often based on the consequences they may have on the way their selves are presented online (self-presentation). Participants considered sharing in social network applications such as Facebook and Twitter as a means to maintain their social capital. In fact, based on the outcomes of the study, social capital appears to be a driver for sharing in social network sites. This resonates with previous research, such as the study of Buckel and Thiesse (2013), as well as the work of Vitak (2012) on self-disclosures in social network sites for relationship maintenance, but also with studies on self-promoting disclosures (e.g. Buffardi and Campbell, 2008; Mehdizadeh, 2010; Ong et al., 2011).

6.7 Conclusion

The research presented in this chapter has focused on the mechanisms that people employ when making disclosure decisions. We conducted a series of focus groups that acted as a follow-up to the survey and focused on exploring how people articulate their location sharing attitudes. Following that, we undertook a detailed qualitative analysis of the data that led to the grouping of all themes into three main categories: social capital, trust in the applications involved, and functionality (i.e. the functional value of their privacy decisions).

Up until this chapter we have explored existing theories of privacy decision-making and their relation to location data; we provided evidence of the privacy paradox on location data along with a deeper analysis of the privacy trade-off. Both of the studies presented in these two chapters (the online survey, and the focus groups) have highlighted the

role of context in privacy decisions. Context appears to be a key mediator in privacy decisions and it becomes far more varied and complex when these decisions are made away from the desktop. The following chapter aims to bring together these two studies and develop a theoretical model based on the findings.

Chapter 7

The Isorropic Model of Contextual Privacy Decisions

In an attempt to point out the disconnect that often exists between people's beliefs about the exposure of their data online and the reality that lies behind their actual online behaviour, in Chapter 4 we developed a model called the Distance Model of Belief, Behaviour and Affordance (DMBBA). The model represents the discrepancy between people's privacy attitudes and their actual disclosure behaviour through its "Belief-Behaviour distance". We wished to shed light upon the paradoxical nature of privacy decision-making depicted through this distance. To that purpose, we conducted two studies: an online survey (presented in Chapter 5), and a series of focus groups (in Chapter 6). The studies unravelled the complex nature of privacy decisions and highlighted that context is a key mediator in these decisions.

In this chapter we aim to explain people's privacy decision mechanisms through the development of a conceptual model, called the Isorropic Model. The model brings together the findings from both the survey and focus group studies with the hope to uncover how disclosure decisions arise and the paradox occurs.

7.1 The matrix revisited

The motivation behind our research is to inform the design of privacy systems by examining the underlying mechanisms that people employ when they make privacy decisions. The first stage — reported in Chapter 5 — involved the investigation of the existence of the privacy paradox in location sharing through an online survey. Survey participants were presented with a number of scenarios and in each of them they were asked whether they would share their location and afterwards they were prompted to justify their response. A qualitative analysis of the justifications took place, which were coded in order

to develop a set of themes that point out the reasons behind people's privacy decisions. Following the analysis, all themes were interpreted and mapped together leading to the development of a matrix shown in Chapter 5, Table 5.10. The study showed that the privacy paradox does apply to location data and that although people state that they care about their privacy, in practice their decisions are heavily influenced by contextual factors.

The study that followed the survey involved a series of focus groups — presented in Chapter 6 — which is a useful method for eliciting detailed information on the subject. Three separate focus groups took place, followed by a qualitative analysis of the data during which a number of themes emerged. The different themes were then grouped together into three separate categories — *social capital*, *trust*, and *functionality* — and can be found in Table 7.1.

Identified Themes from Focus Groups	
Category	**Themes**
Social Capital	Self-presentation, Interacting with others, Who sees my location, Interesting location (out of the ordinary), Apathy (not interested in location sharing)
Trust	Debating trust/ Not sure, Trust in the app, Not considering trust, Rational Ignorance, Valence effect, Lack of control, Concerns over data, Priv. settings issues, Suggestions for priv. settings
Functionality	Application Benefits, Benefits in this scenario, Costs in scenario, Safety issues, Implicit Location in Twitter (not geo-tagged)

TABLE 7.1: Themes developed during the analysis of the focus groups.

As the survey and focus group tables illustrate, many themes were common in the survey data and the focus group data. For instance, a common theme was *Helpful Application*, which showed that in the cases that an application provides a clear and usable benefit, many participants would be willing to share their location with this application. However, as the quantity of the survey data was far greater, the themes derived from the survey are much more numerous than the themes derived from the focus groups. Yet, the focus-group themes provided much greater depth, because the participants were given the opportunity to discuss in detail their answer in each scenario presented to them.

An important finding from the focus groups was that data sharing is often a tool employed by participants in order to maintain their social capital. A theme related to this is *Self-presentation*, which shows that people share their location as a means of impression management — as some undergraduate participants admitted for "showing off". Part

of this impression management process is also a theme called *Interesting location*, which highlights that participants wished to share their data when they were at a location of interest, not at home and not at a location where they go on an every day basis.

Another difference between the findings of the two studies is that in comparison with many of the survey participants and the research students, who cared about the affordances of their data from the application side (e.g. themes *Lack of trust in the application*, and *Concerns over data*), the majority of the undergraduate students who took part in the focus groups actively ignored this aspect. As described in the previous chapter, these participants showed signs of the valence effect and rational ignorance.

Grouping into Final Categories	
Category	**Themes**
Social Capital	Self-presentation, Interacting with others, *Location visible to others*, Interesting location - out of the ordinary, Who sees my location, *Sharing experience with others*, Apathy - not interested in location sharing, *Willing to share with friends only*
Trust	Debating trust/Not sure, Trust in the application, No consideration about trust, Rational Ignorance, Valence effect, *Lack of trust in the application*, *No data manipulation by the application*, Concerns over data (*Data useful only to companies*), Lack of control, Priv. settings issues, Suggestions for priv. settings, *Wishing control over data*
Functionality	Application Benefits (*Helpful*, *Convenient*, *Speed of access*), Benefits in specific scenario, Benefits in specific scenario, Costs in scenario, *Weighing the benefits versus costs in this situation*, *Privacy Reasons*, Safety issues, *Sensitive Locations*, *Location not visible to others* (as a positive aspect), *Existence of alternative options*, *No benefits from location sharing*, *Unwilling to share location*

TABLE 7.2: Grouping the themes from both studies into the categories. Themes relevant solely to the survey are presented in italics.

At this stage, having conducted two separate studies that both generated a number of themes related to the factors that influence people's privacy decisions, it appeared indispensable to revisit matrix 5.10 and map the identified themes of the survey against the themes that were developed in the focus groups.

By combining the themes of the focus groups along with the themes of the survey we developed a combined matrix of all the contextual and situational factors that influence people's decision-making processes (Table 7.2). The themes were grouped together into the three main categories that emerged in the analysis of the focus groups; *social capital,*

trust, and functionality. Themes that are relevant solely to the survey are presented in italics.

Social Capital. Social capital is a concept that deals with the benefits that arise from people's social interactions; therefore this category encompasses all the themes that help people manage their social capital. Managing different audiences that would access their information or managing the ways through which their self would be presented to these audiences was a significant part of this category (e.g. the themes *Who sees my location*, *Self-presentation*). Similarly to our study, Palen and Dourish (2003) distinguished the *identity boundary* as one of their three privacy boundaries, which reflects on how people manage their self-presentation in different audiences. Our findings are also in agreement with the work of Patil et al. (2012); in a study focusing on location-sharing services they showed that people share their location as a means of self-presentation, but also as a means to connect with social and professional circles.

Trust. Trust in online applications was a common topic in both studies that we presented in the previous chapters. A number of different themes related to trust (either as lack of trust, positive trust, or indifference towards trust as a subject altogether) and concerns arising from lack of trust (concerns over people's data, or the privacy settings used) emerged during the studies.

Functionality. The third category, functionality, defined earlier as the quality of having a practical use, refers to all cases where there is a practical gain from trading one's information, usually in the form of the functionality of an application. A variety of themes were included in this category, all of which dealt with either benefits or costs in people's disclosure decisions (e.g. *Application Benefits*, *Safety Issues*, and *Existence of alternative options*).

In the analysis of the survey data a few situational factors were identified as themes of their own (such as public event, private event, work-related event), however during the process of mapping these themes along with the themes from the focus groups it became apparent that all these themes belong to a greater category, which is *context*. Therefore, Table 7.2 does not explicitly contain a theme about context, but this is because it was a key factor in all themes.

Grouping all the study data into these categories of privacy decision-making does not mean that the issue of online privacy management has now been simplified. Instead, each of them represents a group of themes; hence it includes a plethora of different factors that affect privacy decision-making. In fact, the existence of so many different factors poses significant challenges for the design of privacy systems. This matter is discussed further in another section of this chapter.

Through the interpretation of the outcomes of the analysis, we developed a theoretical model for contextual privacy decision-making, which we present in detail in the following section.

7.2 The Isorropic Model

Based on the qualitative analysis from the survey and the focus groups we developed the model presented in Figure 7.1. The model builds on the three main categories that arose from these studies: *trust*, *social capital*, and *functionality*. It is depicted as a scale that represents the trade-off process with its two ends representing the costs and the benefits of a privacy decision. We call it the Isorropic Model for privacy decision-making — from the Greek word "isorropia" that means balance — as it represents the cost-benefit evaluation that is often stressed in related research. However, we go beyond previous work in de-constructing the mechanisms involved in that balance. At the centre of the scale the three factors of the model are displayed: trust, social capital, and functionality. Depending on their values these can become weighted as either costs or benefits. At all times the balance of the scale is eventually controlled by the balancing point, which conceptually can be thought of as the *context*. This happens because each of the three factors is largely affected by context. In that sense the balance of the scale depends on the weighting of the three factors for a given decision, and the context in which that decision is being made.

The key idea captured by the Isorropic model is that privacy decisions can be factored into elements of trust, social capital and functionality (each of which may be positive or negative), and that all three are mitigated by the context of the decision. The Isorropic Model is an entirely reflexive model, in the sense that the factors that affect people's privacy decisions are entirely personal; these are based on the fact that the benefits accrue to the individual solely. Naturally, it can be argued that in practice privacy is not only a private good but also a public good and it can affect not only individuals but society as a whole (O'Hara, 2010). However, this research explores privacy decisions from the viewpoint of the individual with a focus on their perceived costs and benefits to the individual (not the society).

7.2.1 The role of context in privacy decision-making

Context as a factor for privacy decision-making is related to all three factors of the model — social capital, trust in the application, and functionality. Ultimately, the balance of the scale of the Isorropic Model (Figure 7.1) is controlled by context. The three factors can move about on the scale depending on the scenario. That means that this model does not regard social capital, trust and functionality either as benefits or costs at all

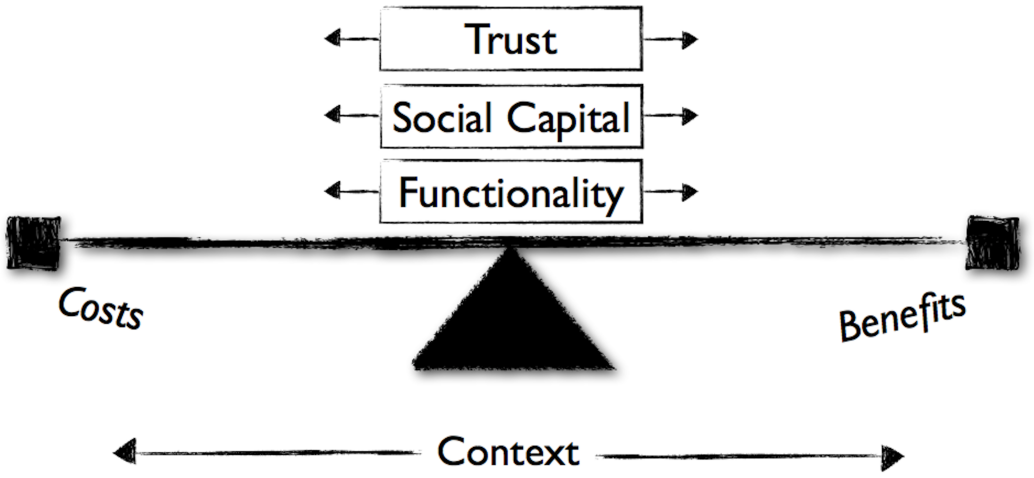


FIGURE 7.1: The Isorropic Model.

times. Each of them can be seen as positive or negative depending on the specific scenario. For instance, a person may wish to share their location with their Facebook friends when being on holidays, but not when being at a local cafe, because it will not attract the interest of their friends. For a given person in a given scenario the three factors (i.e. trust, social capital and functionality) resolve into a position upon the scale of the model, but the final decision may be different depending on the context in which that scenario plays out. That also means that while context is key to all three factors, it is also independent, which is why we have depicted it as the pivoting point of the Isorropic Model.

The fundamental role of context has also been highlighted by several scholars, as described earlier in Chapter 4. Nissenbaum (2010) argues for "the right to *appropriate* flow of personal information" and introduces the theory of *contextual integrity* and its constituents: social context, and context-relative informational norms. Both are dependent on the types of information in the given context, the roles of all the actors in this information exchange and the principles under which this exchange takes place. When these norms are broken, privacy is violated. Similarly to our approach to privacy, Barkhuus (2012) argues that context needs to be taken into account by privacy technologies and employs Nissenbaum's theory of contextual integrity in practice. More specifically, she uses empirical data on personal information management in order to explain contextual integrity online. Grudin (2001) talks about a "steady erosion of a clearly situated action" while he argues that people lose control over their privacy due to the ability of digital technologies to store contextual information indefinitely. Significantly, Palen

and Dourish (2003) argue that the privacy boundaries move dynamically as the context changes. This is in complete agreement with the role of context in the Isorropic Model, because as the context changes all three factors — social capital, trust, and functionality — change the way in which they weigh on the decision.

The following subsection presents a case study aiming to illustrate how privacy decisions are depicted by the Isorropic Model.

7.2.2 Case Study

This is a case study of a participant, who took part in the focus groups presented in the previous chapter. The purpose is to depict the underlying reasoning mechanisms of a participant through the Isorropic Model. The pseudonym 'Jane' is used to refer to the participant.

Scenario "Wikipedia"
In this scenario participants were asked whether they would share their location with Wikipedia in order to explore interesting things around the campus of the university. Jane pointed out that she does not trust how private Wikipedia is. She also stated that there are alternative services online she can use instead of Wikipedia in order to retrieve location-based information (Google Maps was used as an example). On the other hand, the fact that many people use Wikipedia was a positive factor for the participant. Therefore, her answer in the scenario-based question (the question was *"Would you allow Wikipedia to determine your location?"*) was "Maybe".

Scenario "Facebook"
In the Facebook scenario participants were asked whether they would share their location if they were at the airport about to go on holidays. Jane was eager to share her location, hence her response in this scenario was "Yes". Sharing her location offers Jane a means to present herself to other people and attract their attention. In that sense, she stated that she would share her information if she were at an interesting location (e.g. on holidays). She particularly cared about her social privacy, as she commented on cases when she did not wish a friend or her mother to know her location. She pointed out that as long as her privacy settings are set to private, other people will not have access to her information. In that sense, she also noted that in the past Facebook has changed the privacy settings on various occasions resulting her settings not to be set to private any more. However, this did not have any serious effects on her trust levels towards Facebook, offering signs of a biased behaviour, such as rational ignorance. In Jane's own words:

"You think, because it is such a big, like expensive thing, it must regulated quite a lot. You think that, like you know, if there was anything dodgy going on it would have been hopefully found out by now."

Jane's decisions depicted by the Isorropic Model

Based on the information provided by Jane we describe how the Isorropic Model depicts her reasoning mechanisms in both cases (Wikipedia and Facebook).

In the case of Wikipedia, *functionality* was a factor that was evidently perceived as a cost by Jane. Similarly, *trust* to Wikipedia was a negative factor for her, yet the fact that it is a popular service turned it into a less negative one. In the specific scenario that was presented to Jane her answer was "Maybe", as she was not certain whether she would share or not her location. Therefore in the given *context* the balance of the Isorropic Model, as shown in Figure 7.1, did not move towards a specific direction.

In the second scenario, *functionality* did not appear as a factor in her reasoning, since there was no practical outcome expected from her location sharing on Facebook. However, depending on the *context* there could be a positive or a negative outcome to her *social capital*. More specifically, in case she were at an exciting location she would want to share it with her friends. Yet, in case she were at an ordinary location (e.g. at the university) or she did not wish specific people to see her whereabouts, the effect on her social capital would be negative. In addition to this, she did not show any signs of lack of *trust* towards Facebook. In the scenario presented to Jane her answer was positive, meaning that she was willing to share her location on Facebook. Therefore, in this context the scale of the Isorropic Model (Figure 7.1) would move towards the benefits end.

It should be noted, however, that as *context* ultimately controls the balance of the scale in the Isorropic Model, each sharing decision is highly-context dependent. This implies that if the participants were presented with different scenarios, their answers could easily be different and eventually the balance of the model would change.

7.2.3 Comparing our model with others

Prior studies in the privacy trade-off research area have presented (and some tested) theoretical models of the trade-off. Table 7.3 includes some representative models of latest research and highlights key commonalities and differences to our model. Most of the models make use of the two main characteristics of the trade-off; distinguishing between costs and benefits from disclosure. In a similar way to our research, Krasnova et al. (2010) made a clear distinction in their model between perceived risks and perceived benefits. Our work confirms their findings, however it adds a few significant factors for disclosure decisions as it addresses the privacy trade-off directly. The balance of the Isorropic Model depends on context, as it can easily affect all three factors of the model (trust, social capital, functionality). In addition to this, our model focuses explicitly on the impact perceived social capital may have on disclosure decisions.

Examples of other studies				
Paper	**Focus of Study**	**Method**	**Model**	**Key Similarities/ Differences**
Krasnova et al. (2010)	People's motivations to disclose data about themselves.	Model developed based on focus groups and later tested through a survey.	Structural Equation Model of self-disclosure, consisting of two main parts, perceived privacy risks and perceived benefits.	Context not part of this model. Focus on perceived costs versus benefits, whereas we focus on the trade-off as a process. Relationship maintenance and enjoyment seen as benefits.
Kehr et al. (2013)	Develop a conceptual model based on background research focusing on general and situated factors of the trade-off but also on bounded rationality issues.	Proposed method is to test the model through a driving-behaviour app that collects various data (e.g. location, driver characteristics).	Model distinguishes general factors from situated ones, perceived risks and benefits and the relationships between all of them. Institutional trust seen as a general factor, affective states as a situated one.	Distinguish between general and situated factors in the trade-off and highlight the privacy paradox. Privacy concern is part of the model. Explore the role of emotions and affect on decision-making. Trust also plays an important role in this study.
Li et al. (2011a)	Focus on the impact that initial situated factors may have on people's privacy decisions.	Model developed and tested through an experiment (a vendor's website).	Combining privacy calculus with the stimulus organism response model. Initial emotional reactions to an unknown application and fairness levers affect people's disclosure intentions.	Focus on context in relation to application characteristics and their effect on people's emotional and cognitive reactions. Privacy concern is part of the model.
Li (2012)	A systematic review of 15 well-known theories in privacy research.	Literature review of previous theories.	The dual calculus model (combining the privacy calculus with the risk calculus model). Plus a decision table is to predict disclosure intentions.	Different approach: decision table based on risk levels. The privacy calculus is only a part of the model.
Buckel and Thiesse (2013)	Focus not only on people's behavioural intentions but also their actual behaviour.	Model tested through a survey plus data collected from Facebook profiles.	An extension of the model of Krasnova et al (2010) — adding a new factor ("privacy preferences").	Different approach: focus on the privacy paradox through data collection from Facebook. Privacy preferences are part of the model.

TABLE 7.3: Examples of models presented in other privacy calculus studies.

A key difference in comparison with the rest of the models of privacy in Table 7.3 is that we consider all factors to be potentially beneficial or costly depending on the *context*, whereas other models categorise them under either costs or benefits. For instance, sharing one's location with their Twitter followers when attending a professional event could be seen as beneficial to the participants of our studies; however location sharing when at a private gathering was perceived as negative. In that sense, self-presentation at a public, work-related event was a benefit but when at a private event it was viewed as a cost. Our model aims to depict the privacy trade-off, and in that sense it represents people's actual privacy decision-making. Therefore, privacy attitudes are not represented in the Isorropic Model. Rather these are absorbed into the factors. For example, if an individual falls into Westin's class of privacy fundamentalist (Kumaraguru and Cranor, 2005), then we might assume them to be more distrustful of companies, or to weigh utilitarian gains lightly.

Many studies do not take context into account (for instance the work of Krasnova et al., 2010; Buckel and Thiesse, 2013). Some studies focus on specific situated factors, such as emotions and affect, and their role in the privacy calculus (Buckel and Thiesse, 2013; Li et al., 2011a). For instance, Li et al. (2011a) showed that initial emotions formed from the first interaction with an application along with the presentation of relevant information, later on have an impact on people's privacy decisions. Other research has combined this aspect of the trade-off along with the issue of incomplete information (Kehr et al., 2013). Our study places context at the centre of privacy decisions, as it regards context to play a key role in privacy decision-making. For clarification purposes, in our model context encompasses all the aspects that are parts of the setting of an event, so emotional state, physiological state, temporal aspects, related activities, co-location, presence of other people and so on. We see these as pervasive, and influential across all three factors of our model.

7.2.4 The need for dynamic privacy systems

As new platforms and applications emerge the incentive to share private data will continue to increase. Despite this trend, privacy systems themselves (and the ways in which individuals express their preferences) have changed very little.

There is an on-going debate regarding the privacy settings of the most popular online applications. The main issue is that people's internal privacy preferences do not match the ones offered to them (boyd and Hargittai, 2010; Madejski et al., 2012). In other words, the ways in which systems expect people to express their privacy preferences, which are mainly access-control based, do not match people's actual privacy decision-making.

Several researchers have advocated that access control is inadequate for online privacy management (e.g. Kagal and Abelson, 2010; Mondal et al., 2014). Madejski et al. (2012) showed in practice this mismatch between participants' sharing intentions on Facebook and potential privacy breaches by comparing their reported privacy preferences, their actual privacy settings and their Facebook posts. Another study on Facebook revealed that in photos of people on the network where the settings were modified, only 37% of the time did the settings match the expectations of the participants. As a consequence, the audience of these photos was often larger than expected (Liu et al., 2011). On the other hand, a seven-year long study that collected Facebook profile data (such as hometown, birthdate, contact information, and interests) showed that as time passed the amount of profile data displayed *publicly* on the network decreased (Stutzman et al., 2012a). Still, a study on Google+ showed that 85.7% of the participants — who are active users of this application — occasionally share posts *publicly* (Kairam et al., 2012). This means that the issue of privacy management of people's day-to-day disclosure behaviour in different applications (such as Facebook posts, location check-ins, tweets with location etc.) is rather complex, therefore it requires a more analytical approach.

Recently, researchers began developing tools in order to assist users at grouping audiences and their corresponding privacy settings (e.g. Mazzia et al., 2012; Amershi et al., 2012; Egelman et al., 2011; Lipford et al., 2010, 2008; Reeder et al., 2008; Adu-oppong et al., 2010). van den Berg and Leenes (2010) call this concept *audience segregation*. Based on our own findings, *Location visible to others*, *Willing to share with friends only* and *Who sees my location* were significant factors for disclosure decisions. In addition to this, social network sites themselves started to implement such solutions (e.g. Google+ circles[1], and Facebook lists[2]). However, addressing this issue solves only partially the problem of privacy management. Mondal et al. (2014) also argued that current access control systems are insufficient and developed a privacy model based on exposure (i.e. the set of people expected to discover a piece of information). The model aims to predict and also control exposure while being transparent with the users. Although it has significant advantages, as it focuses on the people who may eventually access a piece of information instead of the people who have direct access to it, it does not explore the potential of the information itself (e.g. its potential for further inferences) and does not focus on the contextual nature of online data sharing. Krishnamurthy (2013) proposes a semantic-based approach in which people take a more active role in privacy management by indicating their desired levels of privacy, before these are automatically translated into privacy settings. As an example to this effort his team began developing a Facebook application (still work-in-progress), which displays visually to the user the audiences of their shared content along with the effects of their past privacy settings. The idea is promising, as it is a transparent privacy method that could potentially raise

[1]Feature of Google+ that lets the users group people into different audience groups (i.e. circles).

[2]Similar feature to the previous one, it has three default audience groups: close friends, acquaintances, and restricted.

user awareness of their privacy behaviour, however it does not offer insight into new privacy mechanisms that would move away from access control.

As already mentioned in Chapter 4, recent approaches to the issue of privacy management are using recommender systems based on machine learning techniques as a potential solution to the issue of privacy management with interesting results, as they offer a more dynamic approach to this issue (e.g. Ghazinour et al., 2013; Li et al., 2011c; Fang and LeFevre, 2010; Li et al., 2011b). Machine learning techniques have the potential to provide simple and useful privacy settings. Still, these approaches have yet to achieve a holistic approach to privacy management, as they strive to offer better privacy settings, but do not take into account *all* the factors that contribute to the contextual nature of privacy decisions. They are mostly based on user profiles, privacy settings history, and community characteristics instead of focusing on the variety of contextual factors that influence people's privacy decisions.

As we discussed in the previous section, the role of context in privacy decision-making is fundamental. As Nissenbaum (2010) suggests, privacy violations take place when the contextual norms of information flow are disrupted. These issues pose challenging questions for researchers and practitioners on how to develop privacy systems that can capture people's actual privacy preferences.

The study presented in Chapter 5 provided evidence that the privacy paradox holds for location data, and sought to (at least partially) explain the paradox by viewing privacy disclosure decisions as part of a process of structuration. The themes identified in this study along with the themes from the study presented in Chapter 6 (as displayed in Table 7.2) were grouped together to form the three main factors of the Isorropic Model and they emphasise the dynamic and contextual nature of privacy decisions. People do not have a standard behaviour towards online sharing; sometimes they wish to share data online and sometimes they do not, depending on the context. Not only are the themes a strong indicator of people's justifications on their privacy decisions, but they also provide insight into the actual factors that influence their decisions.

The Isorropic Model highlights the weakness of current privacy systems to capture people's privacy decision-making mechanisms, as these systems do not take context into account. Privacy systems that rely on static preferences are thus inherently weak and likely to fail as the context of privacy decisions changes. This disconnect between people's sharing mechanisms and the ways in which systems expect them to make decisions poses significant challenges for the design of future privacy systems. The most important question is how we develop *dynamic* privacy systems that take context into account when considering complex factors such as social capital. This is particularly challenging given that the complexity of the social situation/relationships could well be beyond any simplistic attempt to sense or model.

7.3 Conclusion

This chapter has brought together the results of the two main studies of this research aiming to highlight the commonalities and the differences in their results. The survey provided a significant number of different themes around privacy decision-making, whereas the focus groups provided fewer themes but with more depth. As a consequence, the survey findings indicated the plethora of different contextual and situated factors that affect decision making, whereas the focus groups highlighted the primal role social capital plays in people's decision-making, as it is often the main driver for location sharing. Finally, the incorporation of all the themes into the final matrix (displayed in Table 7.2) shows the importance of context in privacy disclosure decisions and stresses the need for more dynamic privacy settings.

Based on these research findings an Isorropic Model for contextual privacy decisions was developed. The key characteristic of this model is that we make no assumptions that any of the three factors (i.e. social capital, trust, functionality) is either a cost or a benefit by default. For a given scenario and individual each of them resolves to either a cost or a benefit (or no value at all) on the scale of the model. However, this can change when the context of a privacy decision changes. Thus we argue that context is pervasive but independent of the three factors; hence it is represented as the pivoting point of the scale rather than as a factor in its own right. This poses a significant challenge for privacy systems, as they are overall static whereas privacy decisions are highly dynamic. There is an increasing need for the development of dynamic privacy systems that reflect the way that people's preferences change as their context changes.

In comparison with related work, decisions in this model are based on a combination of situated and contextual factors. These play a primary role when it comes to location sharing — e.g. being co-located with someone, at a public event, and so on — and they affect all factors of the model (social capital, functionality and trust). Despite the fact that the model shows that these are balancing factors, we observed that self-presentation within the theme of social capital is a positive factor for sharing.

The privacy of individuals online is of pressing concern, and made even more important when considering the spread of location-enabled smart devices. This research has shown that existing theories of privacy decision-making do apply to location and contextual data (we have seen evidence of the privacy paradox, trade-off, rational ignorance and the valance effect), but also that context is a key mediator of privacy decisions. The analysis of the findings of this research along with the Isorropic Model may inform on-going research in privacy management and, hopefully, may assist into the design of future privacy systems to cope with the increasingly contextual privacy decisions.

Chapter 8

Conclusion

> People will continue — sometimes grudgingly — to make trade-offs favouring convenience and perceived immediate gains over privacy; and privacy will be something only the upscale will enjoy.
>
> —Digital Life in 2025, Pew Research Centre 2014

In today's connected world people share a variety of information on the Web; with their friends in social network sites, but also in exchange for services. The advent of smart enabled devices signified the opportunity to exchange such information from any location along with the location information itself. Although these information exchanges provide several benefits, they do raise a number of privacy concerns. People are called to perform a *privacy trade-off*, a cost-benefit evaluation of their sharing decisions. It is suggested that these disclosure decisions are often in discordance with people's privacy attitudes, a concept known as the *privacy paradox*. The focus of this thesis was on the trade-off process and its potentially paradoxical nature, as well as its discrepancy from the actual data collection and usage practices of systems.

In this final chapter we present a summary of the studies we conducted for this thesis, along with the main contributions. The chapter concludes with some directions for future work and final remarks.

8.1 Summary

The first part of this research, presented in Chapter 4, was the development of a framework for analysing location data, which was tested with the analysis of a sample of 32 research-based systems. The study highlighted the power of location data to act as a catalyst in inferring and aggregating other types of information. It also highlighted the role of context as a technical concept that enables systems to make complex inferences

based on location (2nd and 3rd degree). Most importantly, the analysis raised an important issue related to privacy, which is people's lack of awareness regarding the actual affordances of their data. Following this, we developed a conceptual model, called the Distance Model of Belief, Behaviour and Affordance (DMBBA), which brought together the findings of the analysis with the studies that we went on to explore in the rest of this thesis. The model uncovered the issue that is raised by the existence of two "distances": the discrepancy between people's attitudes and actual behaviour (the privacy paradox), and the discrepancy between people's attitudes and the actual affordances of their data.

The next logical step was the design of a study that would explore the existence of the paradox on location data and the underlying reasons behind people's disclosure behaviour. To achieve this, we focused on how people justify their sharing decisions based on the theories of the privacy trade-off and structuration, as described in Chapter 2. The study was presented in Chapter 5 and comprised of an online survey with 150 respondents, which revealed that the paradox does apply to location data and that people's privacy decisions are moderated by contextual and situational factors in a similar manner to the theory of structuration. This chapter also highlighted the role of context as a social concept that involves all the contextual and situational factors that underlie a privacy decision and gives those decisions a dynamic nature.

The analysis of the survey findings led to the design of a follow-up study consisting of a series of focus groups. The follow-up study was presented in Chapter 6 and its purpose was to investigate in more depth the reasons why people decide to make certain privacy decisions and more specifically the role of context in these decisions. The last two studies (the online survey, and the focus groups) highlighted the importance of context in privacy decisions, as context appears to be a key mediator in privacy decision-making.

Finally, based on the outcomes of these studies we developed a conceptual model for privacy decisions, presented in Chapter 7. The Isorropic Model depicts the process of making privacy decisions as cost-benefit evaluations. It takes into account three different factors; *social capital*, *trust*, and *functionality*. For a given scenario and individual, each of them resolves to either a cost or a benefit (or no value at all) during the decision process. The most prominent role is played by *context*, as it controls the balance of the model and ultimately is responsible for people disclosure decisions.

8.2 List of contributions

At this point the contributions of this thesis are briefly listed:

- *A framework for analysis of location data.* Through the development of the framework for analysing location data, which was described in Chapter 4, this thesis highlighted the catalyst role of location data in inferring and aggregating a variety

of other types of information that are often related to the context of the location sharing decision (e.g. related event, co-located people), but may also be unrelated to the context (e.g. health information, social ties). This implies that the exposure of location data in Web applications poses threats to people's privacy, as there is a clear discordance between the impression people have regarding the affordances of their data and the actual affordances of their information through web technologies. The framework that was developed for the analysis of location data can be further employed and extended in other studies that aim to analyse other types of data and their affordances. The analysis of location data in the sample of systems we selected can also be reused and extended. It can be found in Appendix A.

- *A Distance Model of Belief, Behaviour and Affordance (DMBBA).* The disconnect between people's beliefs with regards to their data and the actual affordances of their data was clearly illustrated by the DMBBA model in Chapter 4. The model also highlighted another disconnect that arises from the *privacy paradox*; the discrepancy between people's privacy attitudes and their actual disclosure behaviours. This contribution highlights a gap that currently exists and needs to be addressed by the research community; the discrepancy between people's attitudes and beliefs and the reality of their disclosure decisions.

- *An online survey studying privacy attitudes and sharing decisions.* The survey presented in Chapter 5 addressed the second disconnect that is raised by the DMBBA model with a focus on location sharing. In other words, the survey focused on people's trade-off decisions and the existence of the privacy paradox on location data. The results of quantitative analysis of the survey provided evidence that a paradox between people's privacy attitudes and their actual disclosure decisions does exist, and it applies to location data as well. The qualitative analysis of the survey — in Chapter 5 — found that the theory of structuration can be applied to privacy decision-making. Our findings showed that in practice people tend to negotiate their privacy (free agency), as their privacy decisions are tempered by situation and context (structures). This contribution comprises of the survey questionnaire that was disseminated to the participants (presented in Appendix B), and the qualitative analysis undertaken that resulted in the development of the matrix of themes (shown in Table 5.10).

- *A study comprising of a series of focus groups.* We conducted a series of focus groups — presented in Chapter 6 — aiming to gain a deeper insight into people's privacy decisions and the ways they articulate their justifications on their decisions. The findings of the focus group sessions showed that the research students committed themselves in a trade-off where they thought about the consequences of their decisions and were all well-informed on issues around privacy. In contrast to this, the undergraduate students cared more about who would access their data

and not on privacy issues related to the affordances of their data. The contribution of this study consists of the design of the study, the questionnaires that were given to the participants (presented in Appendix C), and the thematic analysis undertaken.

- *The Isorropic Model of contextual privacy decisions.* The themes developed during the qualitative analysis of the survey as well as the focus groups were grouped together, interpreted and resulted in the design of a conceptual model, called the Isorropic Model. The three main factors of the model are *social capital, trust* in the application, and *functionality* of the application, which represent the categories developed through the interpretation process of the themes. All the themes of both studies also indicated that they are related to a greater category, which is *context*. The research findings point out that privacy decisions are based on a combination of contextual and situational factors. These play a primary role when it comes to location sharing e.g. being co-located with someone at a public event. The contribution here consists of the model itself, along with the analysis of all the themes that we developed in the two studies.

The contributions took place with the hope to address the original hypothesis of this research. This was:

The privacy paradox applies to location data, yet it does not adequately describe the distance between people's beliefs regarding their data and the actual affordances of their data. Privacy disclosure decisions around location can be understood based on discrete factors, including agency and structures, that are mitigated by context.

The research questions that emerged from the hypothesis are the following:

1. What are the affordances of location data and to what extent are people aware of them?
2. How do people perceive and value their location privacy in theory?
3. How do people value their location data in practice during the privacy trade-off?
4. To what extent do people act as agents and to what extent are they influenced by certain structures during the privacy trade-off?
5. Can we develop a model that takes into account the contextual and dynamic nature of location privacy?
6. What are the key factors of this model?

The first question was addressed through the analysis of the sample of technical systems based on the framework we developed, as well as the development of the conceptual model, called DMBBA in Chapter 4. The analysis showed that location data play a primal role in inferring other types of information. Although in the majority of the systems the 1st degree data were exposed with the individual's explicit consent, there was not sufficient information to suggest that the consent of the users was requested before making 2nd and 3rd degree inferences on their data. This implies that the exposure of location data on the Web may cause a number of privacy related issues, as it enables a number of inferences to be made about which people are often unaware. The DMBBA model also highlighted this through the Belief-Affordance distance, which stressed that people are unaware of the actual affordances of their data.

The following three questions — i.e. questions 2, 3, and 4 — were addressed through the design and analysis of the survey, as presented in Chapter 5. The survey outcomes found that people do value their privacy in theory. More specifically, as Figure 5.3 shows approximately 80% of the participants answered that they are concerned about their privacy. Yet, their privacy decisions are paradoxical in comparison with their stated attitudes. As the statistical analysis of the survey showed, there is no strong correlation between people's attitudes towards privacy and their disclosure decisions. This happens because — as the qualitative analysis of the survey showed — in practice their privacy decisions are driven by situation and context (structures). In that sense, they do not act entirely as free agents, but their decisions are tempered by the structures that govern their privacy decisions. The study also suggests that agency most often has a negative influence on people's disclosures, whereas structures tend to have a more positive influence.

The final two questions — i.e. questions 5, and 6 — were addressed by interpreting the outcomes of the qualitative analysis of the survey and focus group data, as well as through the development of the Isorropic Model in Chapter 7. The model was developed based on all the contextual factors that moderate privacy decisions using the themes that emerged in the qualitative analysis of the previous two studies. It consists of three main factors; *social capital*, *trust* in the application, and *functionality* of the application. The model does not regard these three factors either as benefits or costs at all times. Each of them can be seen as a positive or a negative depending on the context. In that sense, each privacy decision is ultimately dependent upon the context.

8.3 Publications

A number of publications came out from the work presented in this thesis.

- A poster paper, entitled *Privacy Implications of Location and Contextual Data on the Social Web*, was presented at the ACM Web Science Conference 2011.

The paper argues for the privacy concerns that are raised from online location sharing that have effects beyond location, since other contextual information can be inferred through location information (Zafeiropoulou et al., 2011).

- The framework for analysis of location data as well as the analysis of the technical systems based on the framework were put together as a book chapter, entitled *Location Data and Privacy: A Framework for Analysis* (Zafeiropoulou et al., 2012).
- A paper, entitled *Unpicking the Privacy Paradox: Can Structuration Theory Help to Explain Location-Based Privacy Decisions?*, was presented as a full paper at the ACM Web Science Conference 2013. The paper presents the outcomes of the quantitative as well as the qualitative analysis of the online survey (Zafeiropoulou et al., 2013).

Finally, a paper that aims to explain people's privacy decision mechanisms and presents the Isorropic Model, entitled *To Share or Not to Share: The Isorropic Model of Contextual Privacy Decisions* has been submitted to a journal and is considered by reviewers.

8.4 Future work

The outcomes of this research call for further extension and improvement. The framework for analysis of location data that we presented in Chapter 4 can be further extended with more properties to unveil different aspects of the analysed information. For example, new properties could address questions such as what are the functionalities/purposes of the inferred information, or more technical questions such as what are the computational costs for 3rd degree inferences and so on. The framework also addresses the issue of transparency that we discussed in Chapters 3 and 4. As soon as people release their data to an application, they lose control over it. However, in a scenario where the data manipulation was completely transparent, people would be held accountable of the data manipulations and the second distance of Figure 4.5 (between people's beliefs and the actual affordances of their data) would not exist. In that sense, the framework may offer a transparent means of uncovering the different aspects of data manipulation.

A further examination of the relationship between structuration theory and privacy decision-making could also provide fruitful insight. As described in Chapter 2, structuration frames behaviour as a balance between structure and agency, where the *structure* refers to the *rules* and *resources* that shape people's behaviour, whereas the *agency* refers to people's ability to act based on their free choices (Giddens, 1984). The findings of the qualitative analysis of the survey in Chapter 5 revealed that privacy decisions can be viewed as a case where the mechanism of structuration theory can apply. According to the "duality of structure", structures are the medium for decision making but at the same time they are also the outcome of the agents' decision-making. By placing "duality

of structure" in the context of online privacy decisions, we could develop a study that would examine in more detail the relationship between structures and agency in privacy decision-making. The outcomes could prove to be useful into enforcing structures (i.e. rules and resources) upon people that would help them make better-informed privacy decisions on the Web.

In Chapter 7 we argued for the importance of developing dynamic privacy systems by presenting the current limitations in online privacy management, reviewing research studies that point to these gaps but also offer potential solutions, as well as by presenting our own conclusions from the studies described in this thesis. As we explained, the main problem is that the ways in which systems expect people to express their privacy preferences, which are mainly access-control based, do not match people's actual privacy decision-making. The Isorropic Model highlights this weakness of current privacy systems to capture people's decision-making mechanisms, as these systems do not take context into account. Therefore, ideally, the next logical step is the development of a contextual system for location privacy management.

Such a system would take into account the highly contextual and dynamic nature of privacy decisions, as well as the user's generic privacy preferences. It needs to focus on the mechanisms that drive people to make certain privacy decisions online. The findings of the focus groups and the survey, as outlined in the previous chapters, can provide the foundation for the model design, as they provide numerous examples of people's justification process for their sharing decisions in various contexts. Not only are the themes a strong indicator of people's justifications on their privacy decisions, but they also provide insight into the actual factors that influence their decisions. In Chapter 4 we reviewed the latest studies in online privacy management discussing their tendency to find more dynamic approaches using machine learning techniques. Taking these background studies into account, the core of the system could be a machine learning algorithm that focuses on context. More specifically, it would use as a basis people's generic privacy preferences, as well as the contextual factors revolving their past privacy decisions as a basis to make better recommendations for their future decisions. In that sense, the overall aim would be to assist individuals into taking better informed privacy decisions by making them aware of their actual disclosure behaviour.

8.5 Final remarks

To conclude, this thesis has focused on unravelling a rather complex topic, which is online privacy, and more specifically it has attempted to shed more light on privacy decision-making. According to the predictions of scientific experts, put together in Pew Research Centre's report on digital life, in 2025 people will continue making privacy decisions favouring immediate gratification and convenience over privacy. In addition to this,

privacy will be "something only the upscale will enjoy" (Anderson and Rainie, 2014). This implies that both distances presented in the DMBBA model — in Chapter 4 — will continue to grow. People will continue to exchange their privacy for immediate benefits, and at the same time systems will take advantage of people's information without their knowledge. The latter means that accountability and transparency mechanisms are now more important than ever to avoid such scenarios. In any case, these future predictions are rather alarming and call for immediate attention.

This thesis raises a number of issues that need to be addressed in order to successfully manage online privacy. First and foremost, a key argument is that all future research directions (as well as the designed privacy systems) on privacy need to place *context* at the centre of their attention. It should be, once again, clarified that when we refer to context, we mean a broad and dynamic concept that is both technical and social and incorporates the plethora of different types of contextual and situational factors under which a privacy decision takes place, but also all the information that can be inferred through a piece of information (in the case of this research that is location data).

Secondly, it offers an extensive analysis of privacy justifications that reveal the plethora of contextual factors that affect people's privacy decisions. The qualitative analysis of the survey and the focus group series provided a deeper insight into people's reasoning behind their privacy decisions, as we analysed a significant number of justifications from the participants' responses. The interpretation stage of these two studies resulted in the development of the Isorropic Model. This model groups all the different factors for decision-making into three main categories *social capital*, *trust*, and *functionality*.

Finally, this thesis points out the failures of current privacy systems, which are mainly static and based on access control. Yet, as privacy decisions are highly dynamic and contextual, it calls for the design of dynamic privacy systems that take the contextual nature of privacy decisions into account. The Isorropic Model provides significant input for the design of new and more sophisticated privacy systems that can be better at understanding the decisions of their users in different contexts.

The issues around privacy on the Web are not going to go away, in fact they are going to become even more evident in our lives and it is vital that we find proper ways to address them at a global level through a collaboration between all the possible stakeholders, including governments, research bodies, and commercial institutions, and with a combination of socio-technical and legal approaches.

Appendix A

Analysis of location data in the sample of systems

Table A.1 contains the list of papers selected for the analysis of location privacy, which was presented in Chapter 4.

Paper	Authors	Conference
1	Sohn et al	Mobile HCI '10
2	Wagner et al	Mobile HCI '10
3	Cherubini et al	MobileHCI '09
4	Robinson et al	MobileHCI '08
5	Von Watzdorf and Michahelles	MobileHCI '09
6	Harper and Taylor	MobileHCI '09
7	Ankolekar et al	MobileHCI '09
8	Yoon et al	MobileHCI '08
9	Preuveneers et al	MobileHCI '08
10	Clawson et al	MobileHCI '08
11	Lovett et al	UbiComp '10
12	Dearman et al	UbiComp '10
13	Lin et al	UbiComp '10

Paper	Authors	Conference
14	Tang et al	UbiComp '10
15	Toch et al	UbiComp '10
16	Cranshaw et al	UbiComp '10
17	Madan et al	UbiComp '10
18	Lim and Dey	UbiComp '09
19	Cui et al	MobileHCI '10
20	Zheng et al	UbiComp '08
21	Stewart et al	UbiComp '08
22	Bamford et al	MobileHCI '08
23	Hang et al	MobileHCI '08
24	Herbst et al	MobileHCI '08
25	Robinson et al	MobileHCI '09
26	Hutter et al	MobileHCI '08
27	Froelich et al	MobileHCI '08
28	You et al	MobileHCI '08
29	Melto et al	MobileHCI '08
30	Anguera and Oliver	MobileHCI '08
31	Brush et al	MobileHCI '10
32	Meschtscherjakov et al	Ubicomp '08

Table A.1: Papers selected for the analysis.

The following part contains the analysis table. The table includes the data categories that were analysed in each system based on the set of properties that are part of the framework for analysis (presented in Table 4.2). Each system that we analysed has a

key number that corresponds to one of the sample papers, for example the first system has the key "PAPER 1", which corresponds to the first paper of Table A.1.

PAPER 1	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data	1st		EC		UF	U	SNS
a. Location name (Location-based Lens)	1st	IPII	EC		UF	U	AFFECTS 0.5
3. People associated with location	2nd	DPII	IC		UF	U	
4. Content associated with location	1st		EC		UF	U	

PAPER 2	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							location-based app
a. Country data	1st	DPII	EC		UF	U	SNS
b. City data	1st	DPII	EC		UF	U	Explicit Access Settings (wh
c. Street data	1st	DPII	EC		UF	U	AFFECTS 1
2. Associated action (activity)	1st	DPII	EC		UF	U	PRIVACY FACTOR
3. Availability level	2nd	IPII	IC		UF	U	Locaccino

PAPER 3	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Phone Picture	1st	DPII	EC	A1C1T1		U	photo /multimodal
2. Location data	1st	IPII		A1C1T1		S	AFFECTS 0
3. Datetime	1st	NPII		A1C1T1		S	
4. UserID	1st	DPII		A1C1T1		S	

PAPER 4	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data	1st	IPII	EC	A1C1T1	U	U	POI
2. Geotagged data (GPS coordinates)	3rd	NPII	EC	A1C1T1	U	S	3rd
3. Points of Interest	1st	NPII		A1C1T1	U	S	AFFECTS 1

PAPER 5	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data (GPS or CellID)	1st	DPII	EC	A1C1T1	U	S	health & environment
2. Location-related risk data	2nd	HPII	EC	T1	U	S	AFFECTS 0.5

PAPER 6	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Photo (Glance)	1st	DPII	EC	A1C1T1	UF	S	photo
2. Location of glance	2nd	DPII	IC	T1	UF	U	Yes/No access toall phone c
3. Activity during glance	2nd	DPII	IC	T1	UF	U	AFFECTS 0
4. Availability status	2nd	DPII	IC	T1	UF	U	

PAPER 7	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							SNS
a. Country data	1st	DPII	EC DD		UF /FOAF	U	access to the context-sharir
b. City data	1st	DPII	EC DD		UF /FOAF	U	plus settings for diff.access
c. GPS-based Street data (GPS coordinates)	1st	DPII	EC DD		UF /FOAF	U	AFFECTS 1
2. Datetime	1st	IPII	EC DD		UF /FOAF	S	PRIVACY FACTOR
3. Personal status	1st	DPII	EC DD		UF /FOAF	U	

PAPER 8	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Photo	1st	DPII	EC	A1C1T1	U Edit	U	photo
2. Location data							AFFECTS 0.5
a. GPS coordinates	1st	DPII	IC	A1C1T1	U	S	
b. Location keyword	1st	IPII	EC	A1C0.5T1	U	U	
3. Datetime	1st	NPII	IC	A1C1T1	U	S	
4. Presence of co-located phones		IPII				S	
5. People/ event keyword	1st	IPII	EC	T1	U	U	

PAPER 9	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							health
a. Current location	1st	DPII	IC	A1C1T1	U	S	3rd
b. Next location	3rd	DPII	IC	A0C1T1	U	S	New
2. Current activity	3rd	DPII	IC	A0.5C1T1	U	S	AFFECTS 0.5
3. Next activity	3rd	DPII	IC	A0C1T1	U	S	
4. Location recognition	3rd	DPII	IC	A0.5C1T1	U	S	
5. Similarity of location,time, activity combination	3rd		IC	A1C1T1	U	S	

PAPER 10	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Photo	1st	DPII	EC	A1C1T1	UF	U	photo
2. Co-location	2nd	DPII	IC	A1C1T1	UF	U	AFFECTS 0
3. Datetime	1st	NPII		A1C1T1	UF	S	

PAPER 11	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							social network & loc.
a. Actual location data	1st	DPII	IC	A1C1T1	UF	S	SNS
b. Calendar-based location data	1st	DPII	IC	A1C1T1	UF	U	3rd
2. Location-related event	1st	DPII	IC	A1C1T1	UF	U	New
3. Timestamp	1st	NPII	IC	A1C1T1	UF	U	AFFECTS 0.5
4. Co-location	2nd	DPII		A0.5C0.5T1		S	
5. Interrelationship between users (online relation	2nd	DPII		A0.5C0.5T1		S	
6. Real-time social event	3rd			A0.5C0.5T1		S	

PAPER 12	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data	2nd	HPII		A1C1T1	U	S	location-based system
2. Set of location-supported activities	3rd	NPII			U	S	3rd
3. List of errand-based locations	3rd	IPII			U	S	AFFECTS 1

PAPER 13	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							location-based system
a. WiFi and GPS coordinates	1st	DPII	EC	A1C1T1	U	S	
b. Location name	3rd	HPII	EC	A0.5C1T1	U	U	3rd
2. Physical distance between user and recipient (s	2nd	DPII	NO			S	
3. Distance from home/ work	2nd	DPII	NO			S	AFFECTS 1
4. Timestamp	1st	NPII	IC	A1C1T1		S	
5. Duration of stay at location	2nd	DPII				S	
6. Frequency of visits	3rd	DPII				S	
7. Number of users having visited the location	3rd	NPII				S	
8. Entropy	3rd	NPII					

PAPER 14	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							
a. Location name (label)	1st	HPII	EC	A1C0.5T1	UF / U / E	U	location-based system
b. GPS coordinates	1st	DPII	EC	A1C1T1	S	S	
c. Actual physical location	3rd	DPII	NO	A0.5C0.5T1	S	S	3rd
2. Location-related activity	1st	IPII	EC	A1C0T1	UF / U	U	AFFECTS 1
3. Location label semantics	2nd	IPII	NO	A1C1T1	S	S	

PAPER 15	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							AFFECTS 1
a. WiFi and GPS coordinates	1st	DPII	EC	A1C1T1	U/ UF / E	S	location-based system
b. Location name (semantic tag)	1st	HPII	EC	A1C0.5T1	U	U	SNS
2. Entropy	3rd	NPII	NO			S	Locaccino
3. Location viewers (who is allowedto view one's l	1st	DPII	EC	A1C1T1	U	U	3rd
4. Users comfort in location sharing	2nd	NPII	NO		S	S	PRIVACY FACTOR

PAPER 16	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							PRIVACY FACTOR
a. GPS coordinates	1st	DPII	EC	A1C1T1	U/ UF / E	S / U	location-based system
2. Entropy	3rd	NPII	NO		S	S	SNS
3. Co-location	2nd	DPII	NO		S	S	
4. Number of co-locations	3rd	DPII	NO		S	S	
5. Social tie between 2 co-located users	3rd	DPII	NO	A0C1T1	S	S	Locaccino
6. Number of social ties within a SNS	3rd	DPII	NO	A0.5C1T1	S	S	AFFECTS 1

PAPER 17	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data	1st	NPII			S	S	NPII (Data are anonymised
2. Co-location	2nd	NPII			S	S	health
3. Physical Proximity Entropy with others	3rd	NPII			S	S	3rd
4. Behaviour changes due to health problems	3rd	NPII			S	S	New
5. User's health status	3rd	NPII			S	S	AFFECTS 0.5

PAPER 18	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data	1st	DPII	IC		U	U	
2. Co-location	2nd	DPII	IC		U	U	location-based system
3. Timestamp	1st	NPII	IC		U	U	3rd
4. Availability status	1st	DPII	IC		U	U	AFFECTS 0.5
5. Reminder (trigger)	3rd	DPII	IC		U	S	2 systems presented
6. Navigation data	2nd	NPII			U	S	

PAPER 19	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data	1st	DPII	EC		UF / U	S	SNS
2. Social event (activity)	2nd	IPII	EC		UF / U	S	AFFECTS 0

PAPER 20	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							AFFECTS 1
a. GPS coordinates	1st	DPII	EC	A1C1T1	S	S	3rd
2. Users' transportation modes (driving, cycling…)	3rd	DPII	EC	A0.5C1T1	S	S	New

PAPER 21	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							
a. GPS coordinates	1st	DPII	EC	A1C1T1	S	S	location-based system
b. Location name (tag)	1st		EC		U	U	POI
2. Location-specific generated content	1st	NPII	EC	A1C1T1	U	U	AFFECTS 1
3. Points of Interest	1st / 2nd	NPII	EC	A1C1T1	U	S / U	

PAPER 22	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							
a. GPS coordinates	1st	DPII		A1C1T1	U	S	3rd
2. Temporal data	1st	NPII		A1C1T1		S	photo
3. Phone Picture	1st	IPII	EC	A1C1T1	U	U	New
4. User polution exposure	3rd	DPII			S	S	AFFECTS 0.5

PAPER 23	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							location-based system
a. Start location	1st	DPII	EC		U	U	AFFECTS 1
b. Destination	1st	DPII	EC		U	U	
2. Location-based task	2nd		EC		U	S	

PAPER 24	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							
a. GPS coordinates	1st	DPII	EC	A1C0.5T1	U	S	mixed reality game
2. Physical proximity to location	2nd	DPII		A0.5C0.5T1	S	S	AFFECTS 1
3. Temporal data	1st	NPII		A0.5C0.5T1	U	S	
4. Co-location data (social presence)	2nd	DPII		A0.5C1T1	U	S	

PAPER 25	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							
a. GPS coordinates	1st	DPII	EC		U	S	POI
b. Target location	1st	DPII	EC		U	U	AFFECTS 1
c. Location marking	1st	DPII	EC		U	U	
2. Route on map data	2nd	DPII	EC			S	
3. Distance from target location	1st	NPII	EC		U	U	
4. Points of Interest	1st	DPII	EC		U	U	

PAPER 26	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							
a. Geolocation data	1st	DPII		A1C1T1	U	S	POI
b. User's current location on map	1st	DPII		A1C1T1	E	S	
2. Location-based Campus information	2nd	NPII		A1C1T1	U	S	3rd
a. Location entropy (room occupation)	3rd	NPII			U	S	AFFECTS 1
b. Presence of people in location	2nd	DPII			U	S	
3. Points of Interest	2nd	NPII		A1C1T1	U	S	
4. Distance from location (POI)	2nd	NPII			U	S	

PAPER 27	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							
a. Geolocation data	1st	DPII		A1C1T1	U	S	POI
2. Points of Interest	2nd	NPII			U	S	
3. User orientation awareness	2nd	IPII		A1C1T1	U	S	3rd
4. Location-based recommendations	2nd	NPII			U	S	AFFECTS 1

PAPER 28	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							
a. GPS coordinates	1st	DPII	EC	A1C1T1	U	S	mixed reality location-base
2. Picture of location	1st	IPII	EC	A1C1T1	U	U	3rd
3. Picture-based location recognisition	3rd	IPII	EC	A0.5C1T1	U	S	AFFECTS 1

PAPER 29	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data	1st	DPII	EC	A1C1T1	U	U	multimodal app.: multitap
2. Public transport data	3rd	NPII	EC	A0.5C1T1	U	S	3rd
							AFFECTS 1

PAPER 30	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Phone Picture	1st	DPII	EC	A1C1T1	U	U	photo
2. Location data (of picture)	1st	DPII	EC		U	U	multimodal app
							AFFECTS 0

PAPER 31	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							
a. GPS coordinates	1st	DPII	EC		U	S	POI
b. Map data	1st	DPII	EC		U	S	AFFECTS 1
2. Phone Picture	1st	DPII	EC		U	U	
3. Activity	2nd	DPII	EC		U	S	
4. User trail	2nd	DPII	EC		U	S	
5. Navigation data	2nd	DPII	EC		U	S	

PAPER 32	Degree Level	Personal Identifiable Information	User knowledge	Data quality	Data Access	Data Source	Comments
1. Location data							AFFECTS 1
a. Product location	1st	NPII	EC	A1C1T1	U	S	POI
b. User location	1st	DPII	EC	A1C1T1	U	S	
2. User activity	1st	DPII	EC	A1C1T1	U	S	
No. of Data Categories =	**196**						
No. of Papers =	32						

Abbreviations explained:

User knowledge/consent.

EC: Explicit Consent, IC: Implicit Consent

Data Quality (values 0 to 1).

A: Accurate, C: Complete, T: Timely

Data Access/Data Source.

U: User, S: System, UF: User Friend

Appendix B

The survey questionnaire

We are conducting a survey on people's online privacy attitudes and particularly their location sharing attitudes. If you are a smartphone or tablet user we would really appreciate your participation in this survey.

The survey will take you about 5-10 minutes. Your identity will remain anonymous; your answers are confidential and will be added for statistical analysis with the answers from other people we are surveying. You may withdraw at any stage of this survey without saving your answers. (Ethics reference number: 1521).

If you want to take part in the prize draw to win one of the three 20 Amazon vouchers please leave your email address at the end of the survey. For further enquiries please contact us at az4g09@ecs.soton.ac.uk Thank you very much for your collaboration!

☐ Please tick (check) this box to indicate that you consent to taking part in this survey.

Section 1. Demographics

1. Are you Male or Female?

- Male
- Female

2. What is your age?

- 18-25
- 26-34
- 35-43
- 44-52
- 53-61
- 62-on

3. Where do you permanently live in?
Choices are presented in a drop-down list from ISO 3166.

4. What is your country of origin?
Same choices as the previous question.

Section 2. Applications Usage

Which of the following applications do you use in your mobile device?

- A. Wikipedia
- B. Social Networks (Facebook, Google+ etc.)
- C. Movies, Music and Event Planner (IMDb, Flixster, athinorama.gr etc.)
- D. Microblogging Applications (Twitter, Weibo etc.)
- E. Location-based Social Networks (Foursquare, Gowalla etc.)

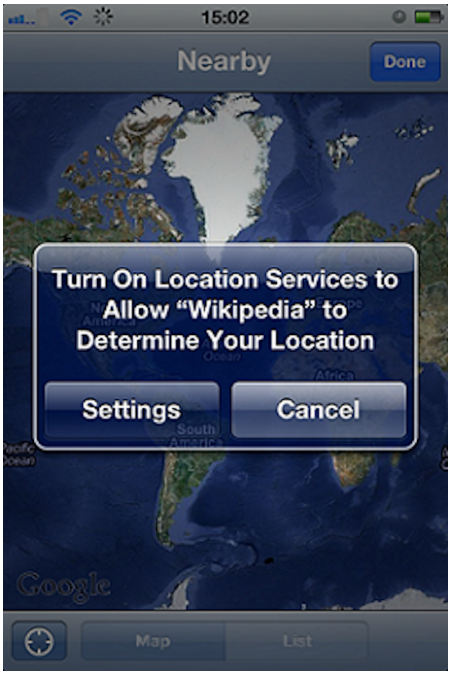
Depending on the applications they choose, participants are directed to answer different scenario-based questions and justify their answers afterwards. The following section shows the scenario questions.

Section 3. Scenarios

A. Wikipedia

Consider the following scenario. You are visiting Berlin for the first time with a friend. While walking around the city centre you wish to explore the museums of the city. Your friend suggests to use Wikipedia on your device and allow it to determine your location. Wikipedia has a "Nearby" feature that shows on a Google map links to Wikipedia pages about locations around you. All the nearby museums of Berlin will appear on the map and through their wikipedia page you may decide which one you wish to visit. Would you allow Wikipedia to determine your location?

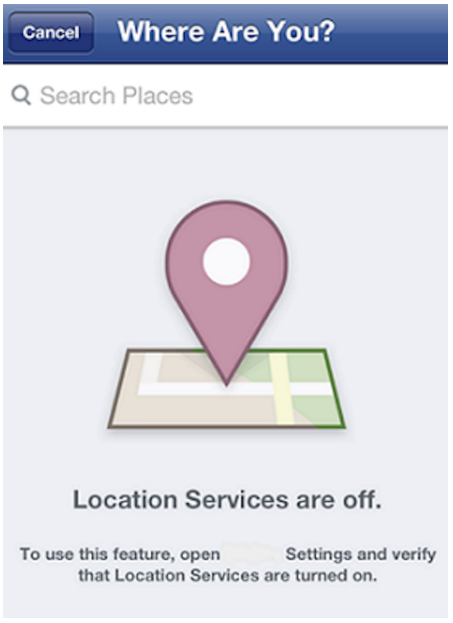
- Yes
- Maybe
- No



B. Facebook

Consider the following scenario. You are visiting a friend (who is also your Facebook friend) in another city. You are going to dinner in a very popular restaurant of that city. Would you post your location on your Facebook wall?

- Yes
- Maybe
- No



C. Foursquare

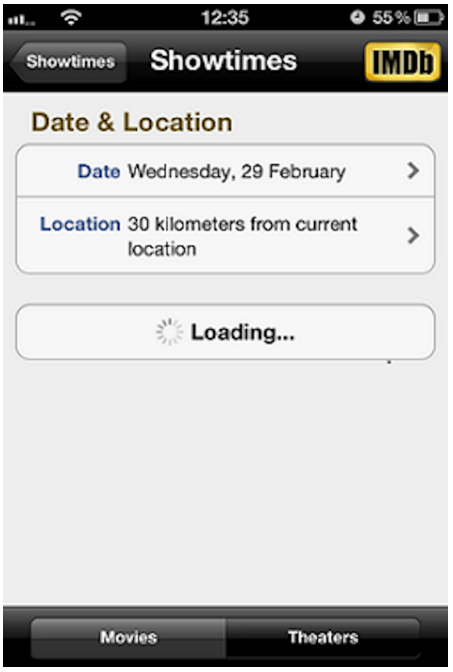
Consider the following scenario. You are at a train station of your city and wish to eat a hamburger first. You want to check if there are any good choices of fast-food restaurants in the area. Would you check-in on Foursquare (i.e. publish your location) to find a nearby restaurant recommended by other Foursquare users?

- Yes
- Maybe
- No

D. IMDb

Consider the following scenario. It's Saturday night, you are out and want to go to the movies but you don't know any cinemas in your area. You may use IMDb to see all the cinemas in the nearby area plus the different showtimes given that you allow IMDb to determine your location. Would you allow IMDb to determine your location?

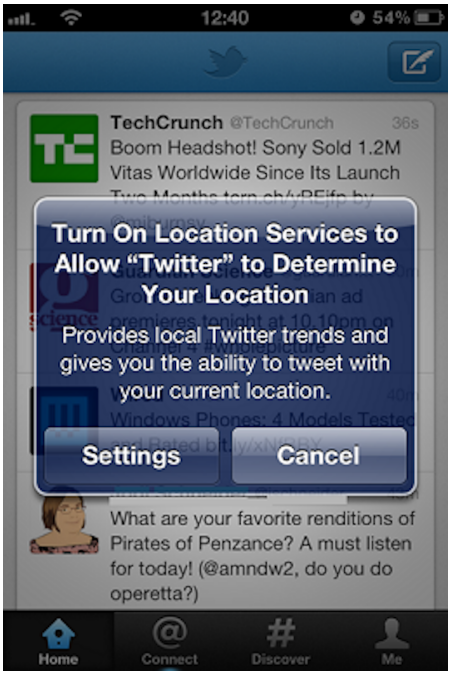
- Yes
- Maybe
- No



E. Twitter

Twitter's Tweet With Your Location feature allows users to selectively add location information to their Tweets. Now please consider the following scenario. You are attending a big public event, which takes place at your city and want to tweet about it. Would you tweet with your location?

- Yes
- Maybe
- No



Section 4. Scenario Justification

At that point participants are requested to justify their responses to the scenario-based questions they have answered. Only the justification questions for the scenarios that have been answered appear on the screen of the participants.

A. Wikipedia

Your answer to the "Wikipedia" scenario was: Yes/ Maybe/ No

Please explain why you gave this answer. For your own convenience, you may use keywords or short phrases in your explanation.

B. Facebook

Your answer to the "Facebook" scenario was: Yes/ Maybe/ No

Please explain why you gave this answer. For your own convenience, you may use keywords or short phrases in your explanation.

C. Foursquare

Your answer to the "Foursquare" scenario was: Yes/ Maybe/ No

Please explain why you gave this answer. For your own convenience, you may use keywords or short phrases in your explanation.

D. IMDb

Your answer to the "IMDb" scenario was: Yes/ Maybe/ No

Please explain why you gave this answer. For your own convenience, you may use keywords or short phrases in your explanation.

E. Twitter

Your answer to the "Foursquare" scenario was: Yes/ Maybe/ No

Please explain why you gave this answer. For your own convenience, you may use keywords or short phrases in your explanation.

Section 5. Privacy Attitudes

The following questions deal with the way you control your privacy online. The image below illustrates the Facebook privacy settings as an example of privacy control. For each of the following indicate whether you agree or not.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I believe I am able to take the appropriate steps to control when and how my location is released online.	○	○	○	○	○
When an application requests my location I am fully aware of the reasons why.	○	○	○	○	○

Now, for each of the following activities identify how often you do them. The image below illustrates the location settings on iPhone and Android phones.

	Never	Not sure	Rarely	Once a month or so	More than once a week	All the time
Do you make use of the privacy settings offered by Web applications to control access to your data?	○	○	○	○	○	○
In your mobile device do you ever have the location services setting on?	○	○	○	○	○	○
How often do you post your location in a Social Networking application? (Facebook, Twitter etc.)	○	○	○	○	○	○
Do you ever allow an application to determine your current location?	○	○	○	○	○	○

For each of the following indicate if it is a concern to you or not.

	Very Concerned	Somewhat Concerned	Not very concerned	Not concerned at all	Don't know
How concerned are you about threats to your online privacy?	○	○	○	○	○
How concerned are you about the fact that your location might be used for other purposes too?	○	○	○	○	○

Location-based websites may use your location to make assumptions about you. Please indicate how important it is for you to control these assumptions for each of the following types of information:

	Very important	Important	Neutral	Not important	Completely unimportant
Your current location	○	○	○	○	○
Your interests/habits	○	○	○	○	○
Your home address	○	○	○	○	○
Your everyday itinerary	○	○	○	○	○
Friends with you at your current location	○	○	○	○	○
Your current activity	○	○	○	○	○
Your current health state	○	○	○	○	○

Section 6. Email

Would you like to take part in the prize draw for an Amazon voucher?

- Yes
- No

If yes, please enter your email address.

Appendix C

Focus group supplementary material

C.1 The focus group handout

Schedule

15:00-15:10 Welcome
15:10-16:00 Main Discussion
16:00-16:10 Wrap-up

Questions

The following questions deal with the way your privacy is managed online.
For each of them please indicate whether you agree or not (tick in the appropriate cell):

	Agree completely	**Agree**	**Neutral**	**Disagree**	**Disagree completely**
Are you satisfied from the privacy settings offered to you by web applications?					
Most of the broadly used privacy settings are based on the concept of "who has access to see your data". Are you satisfied with them?					

Now, please answer the following questions:

	Never	**Rarely**	**Once a month or so**	**More than once a week**	**All the time**
How often do you post your location in a social network (e.g. Facebook, Twitter, Google+)?					
In your mobile device do you ever turn on the location settings?					

A. Facebook Scenario

You are travelling abroad with a friend of yours (also your Facebook friend) and at the moment you are at the airport.

Would you post your location on your Facebook wall? Tick the box that suits you best.

- Yes
- Maybe
- No

Feel free to add any comments about your choice of answer in the area below.

B. Twitter Scenario

You are attending a concert in London and thinking of tweeting about it.

Would you tweet with your location?

- Yes
- Maybe
- No

Feel free to add any comments about your choice of answer in the area below.

C. Wikipedia Scenario

You are visiting Southampton for the first time with a friend and wish to visits the sights of the area near the university.
Your friend suggests using Wikipedia, as it shows on a map links to the Wiki pages of all the nearby sights.

Would you allow Wikipedia to determine your location?

- Yes
- Maybe
- No

Feel free to add any comments about your choice of answer in the area below.

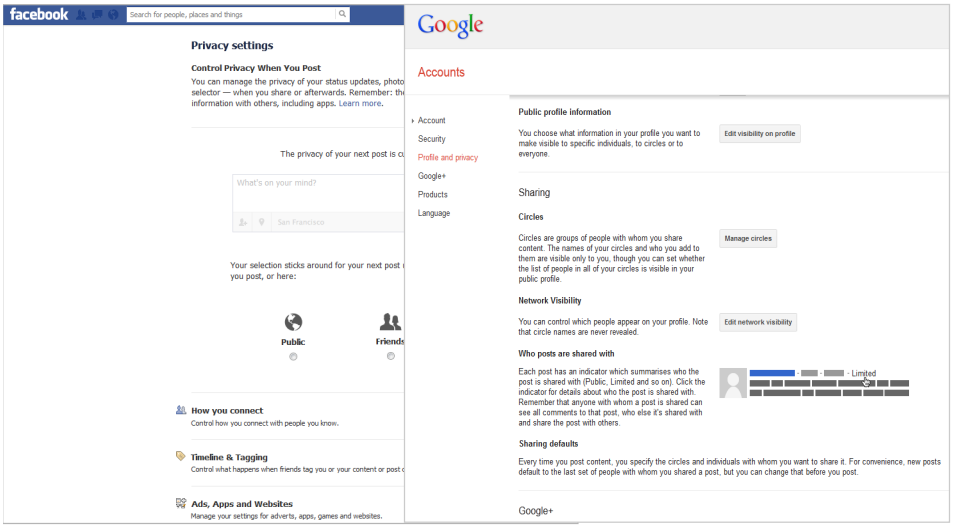
C.2 Presentation

The presentation slides used in the focus group discussion can be found below:

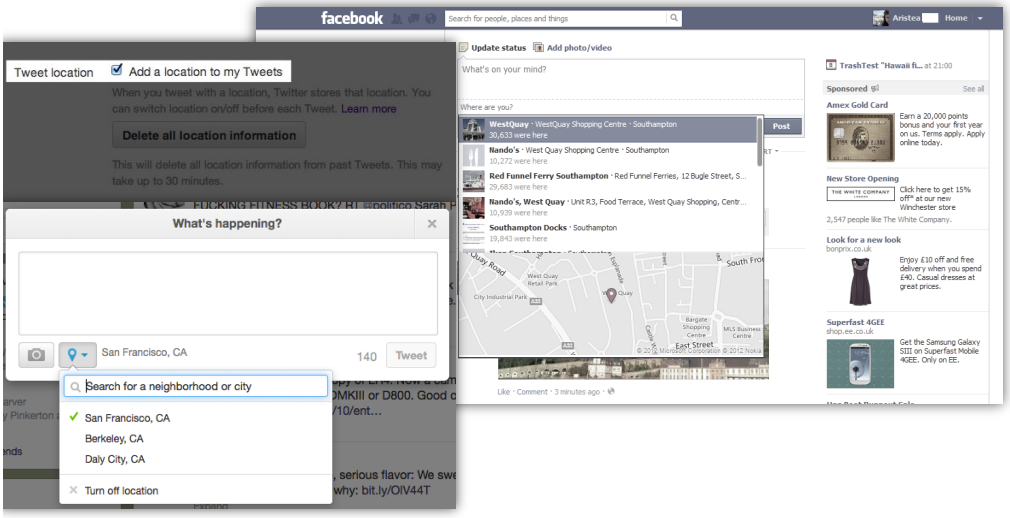
Focus Group Discussion

Moderator: Aristea Zafeiropoulou
Web Science DTC
University of Southampton

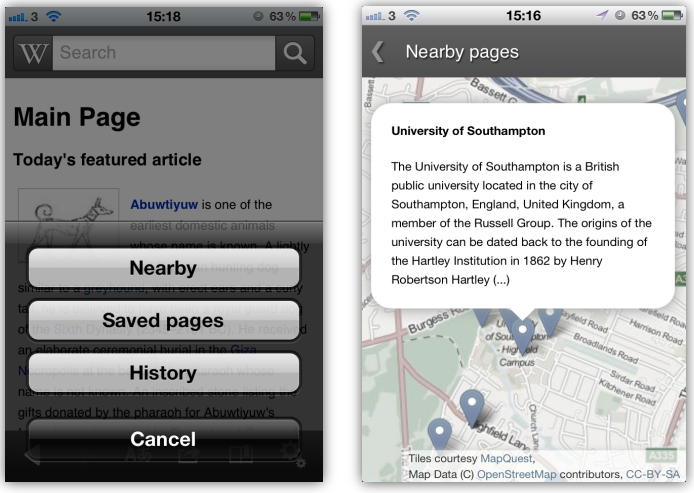
Privacy Settings



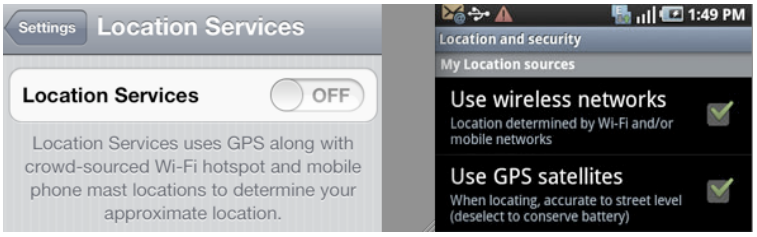
Examples



Wikipedia "Nearby"

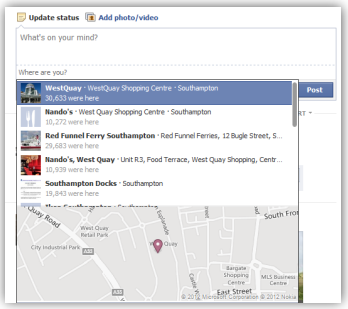


Location Settings



Example of the location settings in iPhone (left) and Android (right)

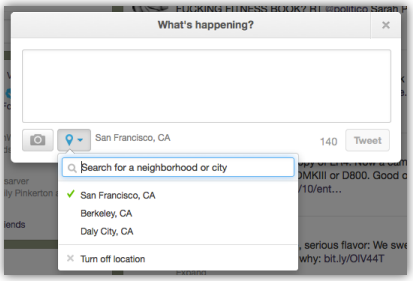
facebook



Scenario “Facebook”

You are travelling abroad with a friend of yours (also your Facebook friend) and at the moment you are at the airport.

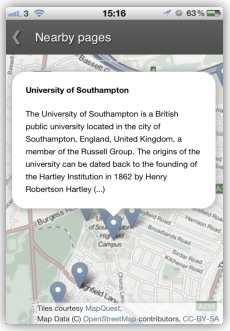
Would you post your location on your Facebook wall?



Scenario “Twitter”

You are travelling abroad with a friend of yours (also your Facebook friend) and at the moment you are at the airport.

Would you post your location on your Facebook wall?



Scenario “Wikipedia”

You are visiting Southampton for the first time with a friend and wish to visits the sights of the area near the university. Your friend suggests to use Wikipedia, as it shows on a map links to the Wiki pages of all the nearby sights.

Would you allow Wikipedia to determine your location?

References

Lyn Y Abramson, Martin E Seligman, and John D Teasdale. Learned helplessness in humans: critique and reformulation. *Journal of abnormal psychology*, 87(1):49–74, March 1978. ISSN 0021-843X.

Alessandro Acquisti. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM Conference on Electronic Commerce*, EC '04, pages 21–29, New York, NY, USA, 2004. ACM. ISBN 1-58113-771-0.

Alessandro Acquisti. Nudging Privacy: The Behavioral Economics of Personal Information. *IEEE Security & Privacy Magazine*, 7(6):82–85, November 2009. ISSN 1540-7993.

Alessandro Acquisti and Ralph Gross. Imagined Communities : Awareness , Information Sharing , and Privacy on the Facebook. *Public Policy*, pages 1–22, 2006.

Alessandro Acquisti and Jens Grossklags. Privacy Attitudes and Privacy Behaviour. Losses, Gains and Hyperbolic Discounting. In Camp J and Lewis R, editors, *The Economics of Information Security*, chapter 1, pages 1–15. Kluwer, 2004.

Alessandro Acquisti and Jens Grossklags. Privacy and rationality in individual decision making. *IEEE Security and Privacy Magazine*, 3(1):24–30, January 2005. ISSN 1540-7993.

Alessandro Acquisti and Jens Grossklags. What Can Behavioral Economics Teach Us About Privacy? In *Digital Privacy: Theory, Technologies and Practices*, pages 363–377. Auerbach Publications, 2007.

Alessandro Acquisti, Leslie John, and George Loewenstein. What is privacy worth? In *Twenty First Workshop on Information Systems and Economics (WISE)*, 2009.

Matthew Adams. Hybridizing Habitus and Reflexivity:: Towards an Understanding of Contemporary Identity? *Sociology*, 40(3):511–528, June 2006. ISSN 0038-0385.

Idris Adjerid, Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. Sleights of privacy: Framing, disclosures, and the limits of transparency. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, pages 9:1–9:11, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2319-2.

Paul S. Adler and Seok-woo Kwon. Social Capital: Prospects for a New Concept. *The Academy of Management Review*, 27(1):17–40, 2002.

Fabeah Adu-oppong, Casey K Gardiner, and Patrick P Tsang. Social Circles: Tackling Privacy in Social Networks. In *Fifth International Conference on Systems and Networks Communications (ICSNC)*, pages 154–159, 2010.

Charu C Aggarwal and Tarek Abdelzaher. Social Sensing. In *Managing and Mining Sensor Data*, chapter 9, pages 237–297. Springer US, 2013.

Irwin Altman. *The environment and social behavior: privacy, personal space, territory, and crowding*. CA: Brooks/Cole, 1975. ISBN 0818501685, 9780818501685.

Saleema Amershi, James Fogarty, and Daniel Weld. Regroup: Interactive machine learning for on-demand group creation in social networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 21–30, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4.

Janna Anderson and Lee Rainie. Digital Life in 2025. Technical Report March, Pew Research Centre, Washington D.C., 2014.

Elliot Aronson. The Theory of Cognitive Dissonance: The Evolution and Vicissitudes of an Idea. In McGarty, editor, *The message of social psychology*, chapter 2, pages 20–35. Oxford: Blackwell, 1997. ISBN 0631197818.

Elliot Aronson and J Merrill Carlsmith. Effect of the severity of threat on the devaluation of forbidden behavior. *The Journal of Abnormal and Social Psychology*, 66(6):584–588, 1963. ISSN 0096-851X.

Rebecca Balebako, Pedro G Leon, Hazim Almuhimedi, Patrick Gage Kelley, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. Nudging Users Towards Privacy on Mobile Devices. In *ACM SIGCHI Conference on Human Factors in Computing Systems*, 2011. ISBN 9781450302685.

Vladimir Barash, Nicolas Ducheneaut, Ellen Isaacs, and Victoria Bellotti. Faceplant: Impression (Mis)management in Facebook Status Updates. In *AAAI Conference on Weblogs and Social Media*, pages 207–210, 2010.

Louise Barkhuus. The mismeasurement of privacy: Using contextual integrity to reconsider privacy in hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 367–376, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4.

Louise Barkhuus, Barry Brown, Marek Bell, Scott Sherwood, Malcolm Hall, and Matthew Chalmers. From awareness to repartee: Sharing location within social groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems, CHI '08, pages 497–506, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-011-1.

Zygmund Bauman. *Thinking Sociologically*. Basil Blackwell Ltd, Oxford, 1990.

Ardion Beldad, Menno de Jong, and Michaël Steehouder. A Comprehensive Theoretical Framework for Personal Information-Related Behaviors on the Internet. *The Information Society*, 27(4):220–232, July 2011. ISSN 0197-2243.

Michael Benisch, Patrick Gage Kelley, Norman Sadeh, and Lorrie Faith Cranor. Capturing Location-Privacy Preferences: Quantifying Accuracy and User-Burden Tradeoffs. *Personal and Ubiquitous Computing*, 15(7):679–694, 2011.

Bettina Berendt, Oliver Günther, and Sarah Spiekermann. Privacy in e-commerce: Stated preferences vs. actual behavior. *Commun. ACM*, 48(4):101–106, April 2005. ISSN 0001-0782.

Alastair R. Beresford, Dorothea Kübler, and Sören Preibusch. Unwillingness to pay for privacy: A field experiment. *Economics Letters*, 117(1):25–27, October 2012. ISSN 01651765.

Michael S. Bernstein, Eytan Bakshy, Moira Burke, and Brian Karrer. Quantifying the invisible audience in social networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 21–30, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1899-0.

Andrew Besmer, Jason Watson, and Heather Richter Lipford. The impact of social navigation on privacy policy configuration. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 7:1–7:10, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0264-7.

James R. Bettman, Mary Frances Luce, and John W. Payne. Constructive Consumer Choice Processes. *Journal of Consumer Research*, 25(3):187–217, December 1998. ISSN 0093-5301.

Grant Blank, Gillian Bolsover, and Elizabeth Dubois. A New Privacy Paradox: Young people and privacy on social network sites. 2014.

Julie Boesen, Jennifer A. Rode, and Clara Mancini. The domestic panopticon: Location tracking in families. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, Ubicomp '10, pages 65–74, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-843-8.

Pierre Bourdieu. *Outline of a Theory of Practice*. Cambridge University Press: London, 1977. ISBN 1107268117, 9781107268111.

Pierre Bourdieu. The forms of capital. In John Richardson, editor, *Handbook of Theory and Research for the Sociology of Education*, pages 241–258. Westport, CT: Greenwood, 1986.

danah m boyd. *Taken Out of Context: American teen sociality in networked publics*. PhD thesis, Cambridge, MA, 2008.

danah m boyd. *it's complicated*. Yale University Press, 2014. ISBN 978-0300166316.

danah m boyd and Eszter Hargittai. Facebook privacy settings: Who cares? *First Monday*, 15(8), 2010.

Ian Brackenbury and Thomas Wong. Online Profile & Reputa1on Perceptions Study. Technical report, 2013.

Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. Misplaced Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science*, August 2012. ISSN 1948-5506.

Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

Alex Braunstein, Laura Granka, and Jessica Staddon. Indirect content privacy surveys: Measuring privacy without asking about it. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, SOUPS '11, pages 15:1–15:14, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0911-0.

Jack W Brehm. Postdecision changes in the desirability of alternative. *Journal of abnormal and social psychology*, 52(3):384–389, 1956.

David Brin. *The Transparent Society*. Basic Books, 1998. ISBN 0738201448.

Thomas Buckel and Frédéric Thiesse. Predicting The Disclosure of Personal Information on Social Networks: An Empirical Investigation. In *Wirtschaftsinformatik Proceedings 2013*, pages 1619–1633, 2013.

David Buckingham. Introducing Identity. *Youth, Identity, and Digital Media, David Buckingham, ed., The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning, The MIT Press, Cambridge, MA, 2008 ; Berkman Center Research Publication*, pages 1–22, 2008.

Laura E Buffardi and W Keith Campbell. Narcissism and social networking Web sites. *Personality & social psychology bulletin*, 34(10):1303–14, October 2008. ISSN 0146-1672.

Thorben Burghardt, Erik Buchmann, Jens Müller, and Klemens Böhm. Understanding user preferences and awareness: Privacy mechanisms in location-based services. In Robert Meersman, Tharam Dillon, and Pilar Herrero, editors, *On the Move to*

Meaningful Internet Systems: OTM 2009, volume 5870 of *Lecture Notes in Computer Science*, pages 304–321. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-05147-0.

Jeffrey A. Burke, D. Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani B. Srivastava. Participatory Sensing. Technical report, Center for Embedded Network Sensing. UCLA: Center for Embedded Network Sensing, 2006.

Moira Burke, Robert Kraut, and Cameron Marlow. Social capital on facebook: Differentiating uses and users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 571–580, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9.

Jacquelyn Burkell, Alexandre Fortier, Lorraine (Lola) Yeung Cheryl Wong, and Jennifer Lynn Simpson. Facebook: public space, or private space? *Information, Communication & Society*, 17(8):974–985, 2014.

L. Jean Camp. Web security and privacy: An american perspective. *The Information Society*, 15(4):249–256, 1999.

Christopher J. Carpenter. Narcissism on Facebook: Self-promotional and anti-social behavior. *Personality and Individual Differences*, 52(4):482–486, March 2012. ISSN 01918869.

Ann Cavoukian. Privacy by Design. The 7 Foundational Principles. Implementation and Mapping of Fair Information Practices. Technical report, Information & Privacy Commissioner, Ontario, Canada, 2010.

Ann Cavoukian. Operationalizing Privacy by Design : A Guide to Implementing Strong Privacy Practices. Technical Report December, Information and Privacy Commissioner, Ontario, Canada, 2012.

Anne S Y Cheung. Rethinking Public Privacy in the Internet Era: A Study of Virtual Persecution by the Internet Crowd. *Journal of Media Law*, 1(2), 2009.

Chao-Min Chiu, Meng-Hsiang Hsu, and Eric T.G. Wang. Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems*, 42(3):1872–1888, December 2006. ISSN 01679236.

Eunjoon Cho, Seth A. Myers, and Jure Leskovec. Friendship and mobility: User movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 1082–1090, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0813-7.

Emily Christofides, Amy Muise, and Serge Desmarais. Hey Mom, What's on Your Facebook? Comparing Facebook Disclosure and Privacy in Adolescents and Adults. *Social Psychological and Personality Science*, 3(1):48–54, May 2011. ISSN 1948-5506.

Barbie Clarke. BFFE (Be Friends Forever): the way in which young adolescents are using social networking sites to maintain friendship and explore identity. In *Proceedings of the WebSci'09: Society On-Line*, 2009.

James S. Coleman. Social Capital in the Creation of Human Capital. *The American Journal of Sociology*, 94:95–120, 1988.

Lizzie Coles-kemp, Yee-lin Lai, and Margaret Ford. Privacy on the Internet: Attitudes and Behaviours. Technical report, VOME Project, 2010.

Alissa Cooper, Deirdre K Mulligan, Henning Schulzrinne, and Erik Wilde. Challenges for the Location-Aware Web. 2010.

Joel Cooper. *Cognitive Dissonance: Fifty Years of a Classic Theory*. SAGE Publications Ltd, 2007. ISBN 1412929733, 9781412929738.

David J Crandall, Lars Backstrom, Dan Cosley, Siddharth Suri, Daniel Huttenlocher, and Jon Kleinberg. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences of the United States of America*, 107(52):22436–41, 2010. ISSN 1091-6490.

Justin Cranshaw, Eran Toch, Jason Hong, Aniket Kittur, and Norman Sadeh. Bridging the gap between physical location and online social networks. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, Ubicomp '10, pages 119–128, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-843-8.

Mary J Culnan and Robert J. Bies. Consumer Privacy: Balancing Economic and Justice Considerations. *Journal of Social Issues*, 59(2):323–342, 2003.

George Danezis and Seda Guerses. A critical review of 10 years of Privacy Technology. In *Proceedings of Surveillance Cultures: A Global Surveillance Society*, pages 1–16, 2010.

Yves-Alexandre de Montjoye, César a Hidalgo, Michel Verleysen, and Vincent D Blondel. Unique in the Crowd: The privacy bounds of human mobility. *Scientific reports*, 3: 1376, January 2013. ISSN 2045-2322.

Bernhard Debatin, Jennette P Lovejoy, Ann-Kathrin Horn, and Brittany N Hughes. Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. *Journal of Computer-Mediated Communication*, 15(1):83–108, October 2009. ISSN 10836101.

Gerardine Desanctis and Marshall Scott. Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory. *Organization Science*, 5(2):121–148, 1994.

Sanorita Dey, Nirupam Roy, Wenyuan Xu, Romit Roy Choudhury, and Srihari Nelakuditi. AccelPrint: Imperfections of Accelerometers Make Smartphones Trackable. In *Network and Distributed System Security Symposium (NDSS)*, pages 23–26, 2014. ISBN 1891562355.

Claudia Diaz and Seda Guerses. Understanding the landscape of privacy technologies. Technical report, Information Security Summit, 2012.

Tamara Dinev and Paul Hart. An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research*, 17(1):61–80, March 2006. ISSN 1047-7047.

Judith S Donath. Identity and deception in the virtual community. In Marc Smith and Peter Kollock, editors, *Communities in Cyberspace*. London: Routledge, 1999.

Paul Dourish. What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8(1):19–30, February 2004. ISSN 1617-4909.

Matt Duckham and Lars Kulik. Location privacy and location-aware computing. In *Dynamic & mobile GIS: investigating change in space and time*, volume 210, page 90. January 2006.

Thomas W Dunfee, N Craig Smith, and William T Jr Ross. Social Contracts and Marketing Ethics. *Journal of Marketing*, 63:14–32, 1999.

Catherine Dwyer, Starr Roxanne Hiltz, and Katia Passerini. Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In *Proceedings of the Thirteenth Americas Conference on Information Systems (AMCIS)*, 2007.

Peter Eckersley. How unique is your web browser? In *Proceedings of the 10th International Conference on Privacy Enhancing Technologies*, PETS'10, pages 1–18, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-14526-4, 978-3-642-14526-1.

Louisa C Egan, Laurie R Santos, and Paul Bloom. The origins of cognitive dissonance: evidence from children and monkeys. *Psychological science*, 18(11):978–83, November 2007. ISSN 0956-7976.

Serge Egelman, Andrew Oates, and Shriram Krishnamurthi. Oops, i did it again: Mitigating repeated access control errors on facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2295–2304, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9.

Serge Egelman, Janice Tsai, Lorrie Faith Cranor, and Alessandro Acquisti. Timing is everything?: The effects of timing and placement of online privacy indicators. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 319–328, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7.

Nicole B Ellison and danah m boyd. Sociality through Social Network Sites. In William H. Dutton, editor, *The Oxford Handbook of Internet Studies*, chapter 8, pages 151–172. Oxford University Press, January 2013. ISBN 9780199589074.

Nicole B Ellison, Charles Steinfield, and Cliff Lampe. The Benefits of Facebook "Friends": Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168, July 2007. ISSN 10836101.

Nicole B Ellison, Charles Steinfield, and Cliff Lampe. Connection strategies: Social capital implications of Facebook-enabled communication practices. *New Media & Society*, 13(6):873–892, January 2011. ISSN 1461-4448.

Lujun Fang and Kristen LeFevre. Privacy wizards for social networking sites. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 351–360, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-799-8.

Reza Farahbakhsh, Xiao Han, Ángel Cuevas, and Noël Crespi. Analysis of publicly disclosed information in facebook profiles. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '13, pages 699–705, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2240-9.

Leon Festinger. *A theory of cognitive dissonance*. Evanstone, IL: Peterson and Company, 1957.

Leon Festinger. Cognitive dissonance. *Scientific American*, 207:93–102, October 1962. ISSN 0036-8733.

Leon Festinger and James M Carlsmith. Cognitive Consequences of Forced Compliance. *Journal of abnormal and social psychology*, 58:203–210, 1959.

Joshua Fogel and Elham Nehmad. Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior*, 25(1):153–160, January 2009. ISSN 07475632.

Michel Foucault. *Discipline and Punish: the Birth of the Prison*. New York: Random House, 1977. ISBN 0679752552, 9780679752554.

Nancy E. Frye and Michele M. Dornisch. When is trust not enough? The role of perceived privacy of communication tools in comfort with self-disclosure. *Computers in Human Behavior*, 26(5):1120–1127, September 2010. ISSN 07475632.

Sarah J. Fusco, Roba Abbas, Katina Michael, and Anas Aloudat. Location-Based Social Networking: Impact on Trust in Relationships. *IEEE Technology and Society Magazine*, 31(2):39–50, 2012. ISSN 0278-0097.

Vaibhav Garg, Apu Kapadia, and L Jean Camp. Peer-produced Privacy Protection. In *IEEE International Symposium on Technology and Society (ISTAS)*, pages 147–154, 2013. ISBN 9781479909292.

Kambiz Ghazinour, Stan Matwin, and Marina Sokolova. YourPrivacyProtector: A Recommender System for Privacy Settings in Social Networks. *International Journal of Security*, 2(4):11–25, 2013.

Anthony Giddens. *The constitution of society: Outline of the theory of structuration.* Cambridge: Polity Press, 1984. ISBN 0520052927, 9780520052925.

Erving Goffman. *The Presentation of Self in Everyday life.* Anchor Books, 1959.

Erving Gofman. *Behavior In Public Places.* The Free Press, 1963.

Joshua Gomez, Travis Pinnick, and Ashkan Soltani. KnowPrivacy. Technical report, UC Berkeley, School of Information, 2009.

Mark Granovetter. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1:201–233, 1983.

Jens Grossklags, South Hall, and Alessandro Acquisti. When 25 Cents is too much : An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information. In *Proceedings of Sixth Workshop on the Economics of Information Security (WEIS)*, 2007.

Jonathan Grudin. Desituating Action: Digital Representation of Context. *Human-Computer Interaction*, 16(2):269–286, December 2001. ISSN 0737-0024.

Harry Halpin and Mischa Tuffield. A Standards-based, Open and Privacy-Aware Social Web. Technical report, W3C Incubator Group, 2010.

Keith N Hampton, Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell. Social networking sites and our lives. Technical report, Pew Research Centre, Washington D.C., 2011.

Il-Horn Hann, Tom S. Lee, Kai-Lung Hui, and I. P. L. Png. Online Information Privacy: Measuring the Cost-benefit Trade-off. In *International Conference on Information Systems*, pages 1–10, 2002.

Eszter Hargittai. Digital Na(t)ives? Variation in Internet Skills and Uses among Members of the "Net Generation". *Sociological Inquiry*, 80(1):92–113, February 2010. ISSN 00380245.

Eszter Hargittai and Eden Litt. New strategies for employment? internet skills and online privacy practices during people's job search. *Security Privacy, IEEE*, 11(3): 38–45, May 2013. ISSN 1540-7993.

Eszter Hargittai and Gina Walejko. The Participation Divide: Content creation and sharing in the digital age. *Information, Communication & Society*, 11(2):239–256, March 2008. ISSN 1369-118X.

Iris Herbst, Anne-Kathrin Braun, Rod McCall, and Wolfgang Broll. Timewarp: Interactive time travel with a mobile mixed reality game. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '08, pages 235–244, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-952-4.

Chris Hoofnagle, Jennifer King, Su Li, and Joseph Turow. How Different are Young Adults from Older Adults When it Comes to Information Privacy Attitudes and Policies? *Available at SSRN 1589864*, 2010.

Mariea Grubbs Hoy and George Milne. Gender Differences in Privacy-Related Measures for Young Adult Facebook Users. *Journal of Interactive Advertising*, 10(2):28–45, 2010.

Kai-Lung Hui, Bernard C. Y. Tan, and Chyan-Yee Goh. Online information disclosure: Motivators and measurements. *ACM Trans. Internet Technol.*, 6(4):415–441, November 2006. ISSN 1533-5399.

Kai-Lung Hui, Hock Hai Teo, and Sang-Yong Tom Lee. The value of privacy assurance: An exploratory field experiment. *MIS Q.*, 31(1):19–33, March 2007. ISSN 0276-7783.

Gordon Hull, Heather Richter Lipford, and Celine Latulipe. Contextual gaps: privacy issues on Facebook. *Ethics and Information Technology*, 13(4):289–302, April 2010. ISSN 1388-1957.

Keise Izuma, Madoka Matsumoto, Kou Murayama, Kazuyuki Samejima, Norihiro Sadato, and Kenji Matsumoto. Neural correlates of cognitive dissonance and choice-induced preference change. *Proceedings of the National Academy of Sciences of the United States of America*, 107(51):22014–9, December 2010. ISSN 1091-6490.

Johanna M Jarcho, Elliot T Berkman, and Matthew D Lieberman. The neural basis of rationalization: cognitive dissonance reduction during decision-making. *Social cognitive and affective neuroscience*, 6(4):460–7, September 2011. ISSN 1749-5024.

Richard Jenkins. *Pierre Bourdieu*. Routledge, 2002. ISBN 0415285275, 9780415285278.

Carlos Jensen and Colin Potts. Privacy policies as decision-making tools: An evaluation of online privacy notices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 471–478, New York, NY, USA, 2004. ACM. ISBN 1-58113-702-8.

Carlos Jensen, Colin Potts, and Christian Jensen. Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies*, 63(1-2):203–227, July 2005. ISSN 10715819.

Leslie K John, Alessandro Acquisti, and George Loewenstein. The Best of Strangers: Context-dependent willingness to divulge personal information. *The Journal of Consumer Research*, 37(5):858–873, 2011.

Maritza Johnson, Serge Egelman, and Steven M. Bellovin. Facebook and privacy: It's complicated. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, pages 9:1–9:15, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1532-6.

Adam N Joinson, David J Houghton, Asimina Vasalou, and Ben L Marder. Digital crowding: Privacy, self-disclosure, and technology. In Sabine Trepte and Leonard Reinecke, editors, *Privacy Online*, pages 33–45. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-21520-9.

Adam N Joinson, Ulf-Dietrich Reips, Tom Buchanan, and Carina B Paine Schofield. Privacy, trust, and self-disclosure online. *Human-Computer Interaction*, 25(1):1–24, 2010.

Chris Jones and Laura Czerniewicz. Describing or debunking? the net generation and digital natives. *Journal of Computer Assisted Learning*, 26(5):317–320, 2010. ISSN 1365-2729.

Chris Jones, Ruslan Ramanau, Simon Cross, and Graham Healing. Net generation or Digital Natives: Is there a distinct new generation entering university? *Computers & Education*, 54(3):722–732, April 2010. ISSN 03601315.

Matthew R Jones and Helena Karsten. Giddens's structuration theory and information systems research. *MIS Quarterly*, 32(1):127–157, March 2008. ISSN 0276-7783.

Lalana Kagal and Hal Abelson. Access Control is an Inadequate Framework for Privacy Protection. In *W3C Workshop on Privacy for Advanced Web APIs*, pages 1–6, 2010.

Sanjay Kairam, Mike Brzozowski, David Huffaker, and Ed Chi. Talking in circles: Selective sharing in google+. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1065–1074, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4.

Dave Karpf. Why Bowl Alone When You Can Flashmob the Bowling Alley?: Implications of the Mobile Web for Online-Offline Reputation Systems. In *Proceedings of the WebSci'09: Society On-Line*, 2009.

Flavius Kehr, Daniel Wentzel, and Peter Mayer. Rethinking the privacy calculus: On the role of dispositional factors and affect. In *Thirty Fourth International Conference on Information Systems*, 2013.

Mark J Keith, Samuel C Thompson, Joanne Hale, Paul Benjamin Lowry, and Chapman Greer. Information disclosure on mobile devices: Re-examining privacy calculus with

actual user behavior. *International Journal of Human-Computer Studies*, 71(12): 1163–1173, December 2013. ISSN 10715819.

Jenny Kitzinger. Introducing focus groups. *British Medical Journal*, 311:299–302, 1995.

Bart P Knijnenburg. Simplifying Privacy Decisions: Towards Interactive and Adaptive Solutions. In *RecSys 2013 Workshop on Human Decision Making in Recommender Systems*, pages 40–41, 2013.

Bart P Knijnenburg and Alfred Kobsa. Making decisions about privacy: Information disclosure in context-aware recommender systems. *ACM Transactions on Intelligent Interactive Systems*, 3(3):20:1–20:23, October 2013a. ISSN 2160-6455.

Bart P Knijnenburg and Alfred Kobsa. Preference-based Location Sharing: Are More Privacy Options Really Better? In *Conference on Human Factors in Computing Systems (ACM SIGCHI)*, 2013b. ISBN 9781450318990.

Bart P Knijnenburg, Alfred Kobsa, and Hongxia Jin. Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies, Special Issue on Privacy Methodologies in HCI*, 0944(404):1–37, 2013.

Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, pages 2–5, March 2013. ISSN 0027-8424.

Apostolos Koutropoulos. Digital Natives: Ten Years After. *MERLOT Journal of Online Learning and Teaching*, 7(4):1–17, 2014.

Hanna Krasnova, Sarah Spiekermann, Ksenia Koroleva, and Thomas Hildebrand. Online Social Networks: Why We Disclose. *Journal of Information Technology*, 25(2):109–125, 2010.

Balachander Krishnamurthy. Privacy and online social networks: can colorless green ideas sleep furiously? *Security Privacy, IEEE*, 11(3):14–20, May 2013. ISSN 1540-7993.

Balachander Krishnamurthy and Craig E Wills. Characterizing privacy in online social networks. In *Proceedings of the First Workshop on Online Social Networks*, WOSN '08, pages 37–42, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-182-8.

Ponnurangam Kumaraguru and Lorrie Faith Cranor. Privacy Indexes: A Survey of Westin's Studies. *ISRI Technical Report*, 2005.

David Laibson. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, (May):443–477, 1997.

Harlan Lebo. The Digital Future Project 2013: Surveying the Digital Future. Technical report, USC Annenberg School Center for the Digital Future, Los Angeles, 2013.

Lawrence Lessig. *Code version 2.0*. Basic Books, New York, 2006. ISBN 0465039146, 9780465039142.

Avner Levin and Patricia Snchez Abril. Two Notions of Privacy Online. *Vanderbilt Journal of Entertainment & Technology Law*, 11:1001–1051, 2009.

Kevin Lewis, Jason Kaufman, and Nicholas Christakis. The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network. *Journal of Computer-Mediated Communication*, 14(1):79–100, October 2008. ISSN 10836101.

Han Li, Rathindra Sarathy, and Heng Xu. Understanding situational online information disclosure as a privacy calculus. *Journal of Computer Information Systems*, 51(1), 2010.

Han Li, Rathindra Sarathy, and Heng Xu. The role of affect and cognition on online consumers' decision to disclose personal information to unfamiliar online vendors. *Decision Support Systems*, 51(3):434–445, June 2011a. ISSN 01679236.

Lei Li, Tong Sun, and Tao Li. Personal social screen - a dynamic privacy assignment system for social sharing in complex social object networks. In *2011 IEEE International Conference on Privacy, Security, Risk and Trust (passat) and 2011 IEEE International Conference on Social Computing (socialcom)*, pages 1403–1408, Oct 2011b.

Nan Li and Guanling Chen. Sharing location in online social networks. *Network, IEEE*, 24(5):20–25, September 2010. ISSN 0890-8044.

Qingrui Li, Juan Li, Hui Wang, and A Ginjala. Semantics-enhanced privacy recommendation for social networking sites. In *2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 226–233, Nov 2011c.

Yuan Li. Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems*, 54(1):471–481, December 2012. ISSN 01679236.

Jialiu Lin, Michael Benisch, Norman Sadeh, Jianwei Niu, Jason Hong, Banghui Lu, and Shaohui Guo. A Comparative Study of Location-Sharing Privacy Preferences in the U.S. and China. *Personal and Ubiquitous Computing*, 13(4):697–711, 2013.

Jialiu Lin, Guang Xiang, Jason I. Hong, and Norman Sadeh. Modeling people's place naming preferences in location sharing. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, Ubicomp '10, pages 75–84, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-843-8.

Nan Lin, Karen Cook, and Ronald S. Burt. *Social Capital: Theory and Research*. Transaction Publishers, 2001. ISBN 0202368947, 9780202368948.

Shi-Woei Lin and Yu-Cheng Liu. The effects of motivations, trust, and privacy concern in social networking. *Service Business*, 6(4):411–424, July 2012. ISSN 1862-8516.

Janne Lindqvist, Justin Cranshaw, Jason Wiese, Jason Hong, and John Zimmerman. I'm the mayor of my house: Examining why people use foursquare - a social-driven location sharing application. pages 2409–2418, 2011.

Heather Richter Lipford, Andrew Besmer, and Jason Watson. Understanding privacy settings in facebook with an audience view. In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, UPSEC'08, pages 2:1–2:8, Berkeley, CA, USA, 2008. USENIX Association.

Heather Richter Lipford, Jason Watson, Michael Whitney, North Carolina, Heather Lipford, Katherine Froiland, and Robert W Reeder. Visual vs. Compact: A Comparison of Privacy Policy Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, pages 1111–1114, 2010. ISBN 9781605589299.

Eden Litt, Erin Spottswood, Jeremy Birnholtz, Jeff Hancock, Madeline E Smith, and Lindsay Reynolds. Awkward Encounters of an Other Kind: Collective Self-Presentation and Face Threat on Facebook. In *ACM conference on Computer supported cooperative work & social computing (CSCW '14)*, pages 449–460, 2014. ISBN 9781450325400.

Ling Liu. Privacy and location anonymization in location-based services. *SIGSPATIAL Special*, 1(2):15–22, July 2009. ISSN 1946-7729.

Yabing Liu, Krishna P. Gummadi, Balachander Krishnamurthy, and Alan Mislove. Analyzing facebook privacy settings: User expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, pages 61–70, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-1013-0.

Sonia Livingstone. Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression. *New Media & Society*, 10(3):393–411, June 2008. ISSN 1461-4448.

Grigorios Loukides and Aris Gkoulalas-Divanis. Privacy challenges and solutions in the social web. *Crossroads*, 16(2):14–18, December 2009. ISSN 15284972.

Anmol Madan, Manuel Cebrian, David Lazer, and Alex Pentland. Social sensing for epidemiological behavior change. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, Ubicomp '10, pages 291–300, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-843-8.

Mary Madden. Privacy management on social media sites. Technical report, Pew Research Centre, Washington D.C., 2012.

Mary Madden, Amanda Lenhart, Sandra Cortesi, Aaron Smith, and Meredith Beaton. Teens, Social Media, and Privacy. Technical report, Pew Research Centre, Washington D.C., 2013.

Mary Madden and Aaron Smith. Reputation Management and Social Media. Technical report, Pew Research Centre, Washington D.C., 2010.

Michelle Madejski, Maritza Johnson, and Steven M Bellovin. A Study of Privacy Settings Errors in an Online Social Network. In *4th International Workshop on Security and Social Networking*, pages 340–345, 2012. ISBN 9781467309073.

Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. Home Location Identification of Twitter Users. 2014.

Delfina Malandrino, Andrea Petta, Vittorio Scarano, Luigi Serra, Raffaele Spinelli, and Balachander Krishnamurthy. Privacy awareness about information leakage: Who knows what about me? In *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society*, WPES '13, pages 279–284, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2485-4.

Miguel Malheiros, Sören Preibusch, and M Angela Sasse. "Fairly truthful": The impact of perceived effort, fairness, relevance, and sensitivity on personal data disclosure. In *Sixth International Conference on Trust & Trustworthy Computing. Lecture Notes in Computer Science, vol. 7904, Springer, Berlin Heidelberg*, pages 250–266, 2013.

Clara Mancini, Yvonne Rogers, Keerthi Thomas, Adam N Joinson, Blaine A Price, Arosha K Bandara, Lukasz Jedrzejczyk, and Bashar Nuseibeh. In the best families: Tracking and relationships. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2419–2428, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9.

Clara Mancini, Keerthi Thomas, Yvonne Rogers, Blaine A Price, Lukazs Jedrzejczyk, Arosha K Bandara, Adam N Joinson, and Bashar Nuseibeh. From spaces to places: Emerging contexts in mobile privacy. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, Ubicomp '09, pages 1–10, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-431-7.

Anoush Margaryan, Allison Littlejohn, and Gabrielle Vojt. Are digital natives a myth or reality? University students use of digital technologies. *Computers & Education*, 56(2):429–440, February 2011. ISSN 03601315.

Stephen T Margulis. Privacy as a social issue and behavioral concept. *Journal of Social Issues*, 59(2):243–261, 2003. ISSN 1540-4560.

Stephen T Margulis. Three Theories of Privacy: An Overview. In Sabine Trepte and Leonard Reinecke, editors, *Privacy online: Perspectives on Privacy and Self-Disclosure in the Social Web*, pages 9–17. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-21520-9.

Alice E Marwick and danah m boyd. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1): 114–133, July 2010. ISSN 1461-4448.

Alice E Marwick, Diego Murgia Diaz, and John Palfrey. Youth, Privacy, and Reputation (Literature Review). *Berkman Center Research Publication No. 2010-5 & Harvard Public Law Working Paper No. 10-29*, 2010.

Alessandra Mazzia, Kristen LeFevre, and Eytan Adar. The pviz comprehension tool for social network privacy settings. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, pages 13:1–13:12, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1532-6.

Robert L McArthur. Reasonable expectations of privacy. *Ethics and Information technology*, 3(2):123–128, 2001.

Aleecia M Mcdonald and Lorrie Faith Cranor. The Cost of Reading Privacy Policies. *Journal of Law and Policy for the Information Society*, 2008.

D Harrison McKnight, Vivek Choudhury, and Charles Kacmar. Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research*, 13(3):334–359, September 2002. ISSN 1047-7047.

Soraya Mehdizadeh. Self-presentation 2.0: narcissism and self-esteem on Facebook. *Cyberpsychology, behavior and social networking*, 13(4):357–64, August 2010. ISSN 2152-2723.

Gustavo S Mesch. Is online trust and trust in social institutions associated with online disclosure of identifiable information online? *Computers in Human Behavior*, 28(4): 1471–1477, July 2012. ISSN 07475632.

Irem Metin and Selin Metin Camgoz. The Advances in the History of Cognitive Dissonance Theory. *International Journal of Humanities and Social Science*, 1(6):131–136, 2011.

Mainack Mondal, Peter Druschel, Krishna P Gummadi, and Alan Mislove. Beyond Access Control: Managing Online Privacy via Exposure. In *Workshop on Usable Security (USEC'14)*, 2014.

David L Morgan. Focus Groups. *Annual Review of Sociology*, 22(1996):129–152, 1996.

Janine Nahapiet and Sumantra Ghoshal. Social capital, intellectual capital, and the organizational advantage. *Academy of management review*, 23(2):242–266, 1998.

Helen Nissenbaum. *Privacy in Context. Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 2010. ISBN 0804772894, 9780804772891.

Helen Nissenbaum. A Contextual Approach to Privacy Online. *Daedalus, the Journal of the Amerixan Academy of Arts & Sciences*, 140(4):32–48, October 2011. ISSN 0011-5266.

Patricia A Norberg, Daniel R Horne, and David A Horne. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs*, 41(1):100–126, 2007.

Amanda Nosko, Eileen Wood, and Seija Molema. All about me: Disclosure in online social networking profiles: The case of FACEBOOK. *Computers in Human Behavior*, 26(3):406–418, May 2010. ISSN 07475632.

Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: universal patterns in human urban mobility. *PloS one*, 7(5):e37027, January 2012. ISSN 1932-6203.

Oded Nov and Sunil Wattal. Social computing privacy concerns: Antecedents and effects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 333–336, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7.

Barack Obama. Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy, 2012.

Andrew Odlyzko. Privacy, economics, and price discrimination on the internet. In *Proceedings of the 5th International Conference on Electronic Commerce*, ICEC '03, pages 355–366, New York, NY, USA, 2003. ACM. ISBN 1-58113-788-5.

Ted O'Donogue and Matthew Rabin. The Economics of Immediate Gratification. *Journal of Beharvioral Decision Making*, 13:223–250, 2000.

Kieron O'Hara. Intimacy 2.0: Privacy Rights and Privacy Responsibilities on the World Wide Web. In *Web Science Conference 2010*, Raleigh, NC, USA, 2010.

Kieron O'Hara and Nigel Shadbolt. *The Spy in the Coffee Machine: The End of Privacy as We Know it*. Oneworld Publications, 2008. ISBN 1851685545, 9781851685547.

Kieron O'Hara and Nigel Shadbolt. Privacy on the data web. *Communications of the ACM*, 53(3):39, March 2010. ISSN 00010782.

Nadia Olivero and Peter Lunt. Privacy versus willingness to disclose in e-commerce exchanges: The effect of risk awareness on the relative role of trust and control. *Journal of Economic Psychology*, 25(2):243–262, April 2004. ISSN 01674870.

Eileen Y L Ong, Rebecca P Ang, Jim C M Ho, Joylynn C Y Lim, Dion H Goh, Chei Sian Lee, and Alton Y K Chua. Narcissism, extraversion and adolescents self-presentation on facebook. *Personality and Individual Differences*, 50(2):180 – 185, 2011. ISSN 0191-8869.

Wanda J Orlikowski. The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, 3(3):398–427, 1992.

Xinru Page, Alfred Kobsa, and Bart P Knijnenburg. Don't Disturb My Circles! Boundary Preservation Is at the Center of Location-Sharing Concerns. In *International Conference on Weblogs and Social Media (ICWSM)*, 2012.

Leysia Palen and Paul Dourish. Unpacking "privacy" for a networked world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 129–136, New York, NY, USA, 2003. ACM. ISBN 1-58113-630-7.

John Palfrey and Urs Gasser. *Born Digital: Understanding the First Generation of Digital Natives*. Basic Books, 2008. ISBN 9780465005154.

Sameer Patil, Greg Norcie, Apu Kapadia, and Adam J Lee. Reasons, rewards, regrets: Privacy considerations in location sharing as an interactive practice. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, pages 5:1–5:15, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1532-6.

Sameer Patil, Xinru Page, and Alfred Kobsa. With a little help from my friends: Can social navigation inform interpersonal privacy preferences? In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, CSCW '11, pages 391–394, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0556-3.

Paul A Pavlou. State of the information privacy literature: Where are we now and where should we go? *MIS Quarterly*, 35(4):977–988, December 2011. ISSN 0276-7783.

B. Piazza-Georgi. The role of human and social capital in growth: extending our understanding. *Cambridge Journal of Economics*, 26(4):461–479, July 2002. ISSN 14643545.

Alejandro Portes. Social Capital: Its Origins and Applications in Modern Sociology. In Eric L. Lesser, editor, *Knowledge and Social Capital*, pages 43–67. Butterworth-Heinemann, Boston, 2000.

Stefanie Pötzsch. Privacy Awareness: A Means to Solve the Privacy Paradox? *The Future of Identity, IFIP AICT*, (216483):226–236, 2009.

Sören Preibusch. Guide to measuring privacy concern: Review of survey and observational instruments. *International Journal of Human-Computer Studies*, 71(12):1133–1143, December 2013. ISSN 10715819.

Marc Prensky. Digital Natives, Digital Immigrants. *On the Horizon*, 9(5):1–6, 2001. ISSN 1074-8121.

Robert D. Putnam. *Bowling Alone: The Collapse and Revival of American Community*. Simon and Schuster, 2000. ISBN 0743203046, 9780743203043.

Daniele Quercia, Diego Las, Casas Jo, Pesce David, Michal Kosinski, Virgilio Almeida, and Jon Crowcroft. Facebook and Privacy: The Balancing Act of Personality, Gender, and Relationship Currency. In *Proceedings of the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012. The AAAI Press 2012*, pages 306–313, 2012.

James Rachels. Why Is Privacy Important. *Philosophy and Public Affairs*, 4(4):323–333, 1975.

Viviane Reding. Your data, your rights: Safeguarding your privacy in a connected world. 2011.

Robert W. Reeder, Lujo Bauer, Lorrie Faith Cranor, Michael K. Reiter, Kelli Bacon, Keisha How, and Heather Strong. Expandable grids for visualizing and authoring computer security policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, page 1473, New York, New York, USA, 2008. ACM Press. ISBN 9781605580111.

Daniele Riboni, Linda Pareschi, and Claudio Bettini. Privacy in location-based applications. chapter Privacy in Georeferenced Context-Aware Services: A Survey, pages 151–172. Springer-Verlag, Berlin, Heidelberg, 2009. ISBN 978-3-642-03510-4.

Chris Rose. The Security Implications Of Ubiquitous Social Media. *International Journal of Management & Information Systems*, 15(1):35–40, 2011.

Carmen Ruiz Vicente, Dario Freni, Claudio Bettini, and Christian S. Jensen. Location-Related Privacy in Geo-Social Networks. *IEEE Internet Computing*, 15(3):20–27, May 2011. ISSN 1089-7801.

Tracii Ryan and Sophia Xenos. Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. *Computers in Human Behavior*, 27(5):1658–1664, September 2011. ISSN 07475632.

Adam Sadilek, Henry Kautz, and Jeffrey P Bigham. Finding your friends and following them to where you are. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 723–732, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-0747-5.

Patricia Sanchez Abril. A (My) Space of One's Own: On Privacy and Online Social Networks. *Northwestern Journal of Technology and Intellectual Property*, 6(1), 2007.

Thomas H. Sander. Social Capital and New Urbanism: Leading a Civic Horse to Water? *National Civic Review*, 91(3):213–234, 2002. ISSN 0027-9013.

Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In *Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications*, WMCSA '94, pages 85–90, Washington, DC, USA, 1994. IEEE Computer Society. ISBN 978-0-7695-3451-0.

A.Allan Schmid. Affinity as social capital: its role in development. *The Journal of Socio-Economics*, 29(2):159–171, January 2000. ISSN 10535357.

Clive Seale. *Researching society and culture*. Sage Publications Limited, 2004. ISBN 0761941975, 9780761941972.

Pan Shi, Heng Xu, and Yunan Chen. Using contextual integrity to examine interpersonal information boundary on social network sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 35–38, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1899-0.

Irina Shklovski, Scott D Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 2347–2356, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1.

Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. The post that wasn't: Exploring self-censorship on facebook. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pages 793–802, New York, NY, USA, 2013a. ACM. ISBN 978-1-4503-1331-5.

Manya Sleeper, Justin Cranshaw, Patrick Gage Kelley, Blase Ur, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. "i read my twitter the next morning and was astonished": A conversational perspective on twitter regrets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 3277–3286, New York, NY, USA, 2013b. ACM. ISBN 978-1-4503-1899-0.

H Jeff Smith, Tamara Dinev, and Heng Xu. Information Privacy Research: An Interdisciplinary Review. *MIS Quarterly*, 35(4):989–1015, 2011.

S. S. Smith and J. Kulynych. It May be Social, But Why is it Capital? The Social Construction of Social Capital and the Politics of Language. *Politics & Society*, 30 (1):149–186, March 2002. ISSN 0032-3292.

Daniel Solove. *The digital person*. New York University Press, 2004. ISBN 0814798462, 9780814798461.

Robert M. Solow. Notes on social capital and economic performance. In Partha Dasgupta and Ismail Serageldin, editors, *Social Capital: A Multifaceted Perspective*, pages 6–10. World Bank Publications, 2001.

Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science (New York, N.Y.)*, 327(5968):1018–21, February 2010. ISSN 1095-9203.

Sarah Spiekermann, Jens Grossklags, and Bettina Berendt. E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior. In *Proceedings of the 3rd ACM Conference on Electronic Commerce*, EC '01, pages 38–47, New York, NY, USA, 2001. ACM. ISBN 1-58113-387-1.

Jessica Staddon, David Huffaker, Larkin Brown, and Aaron Sedley. Are privacy concerns a turn-off?: Engagement and privacy in social networks. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, pages 10:1–10:13, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1532-6.

Charles Steinfield, Nicole B Ellison, and Cliff Lampe. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*, 29(6):434–445, November 2008. ISSN 01933973.

Rob Stones. *Structuration Theory*. Palgrave MacMillan, 2005. ISBN 0333793773, 9780333793770.

Fred Stutzman, Robert Capra, and Jamila Thompson. Factors mediating disclosure in social network sites. *Computers in Human Behavior*, 27(1):590–598, January 2011. ISSN 07475632.

Fred Stutzman, Ralph Gross, and Alessandro Acquisti. Silent Listeners: The Evolution of Privacy and Disclosure on Facebook. *Journal of Privacy and Confidentiality*, 4(2): 7–41, 2012a.

Fred Stutzman, Jessica Vitak, Nicole B Ellison, Rebecca Gray, and Cliff Lampe. Privacy in Interaction: Exploring Disclosure and Social Capital in Facebook. In *International Conference on Weblogs and Social Media (ICWSM)*, 2012b.

Stefano Taddei and Bastianina Contena. Privacy, trust and control: Which relationships with online self-disclosure? *Computers in Human Behavior*, 29(3):821–826, May 2013. ISSN 07475632.

Karen P Tang, Jason Hong, and Dan Siewiorek. The implications of offering more disclosure choices for social location sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 391–394, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4.

Karen P Tang, Jialiu Lin, Jason I Hong, Daniel P Siewiorek, and Norman Sadeh. Rethinking location sharing: Exploring the implications of social-driven vs. purpose-driven location sharing. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, Ubicomp '10, pages 85–94, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-843-8.

Don Tapscott. *Growing up digital: The rise of the Net generation*. New York: McGraw-Hill, 1998. ISBN 0070633614, 9780070633612.

Don Tapscott. *Grown up Digital: How the Net Generation Is Changing Your World*. New York: McGraw-Hill, 2009. ISBN 0071641556, 9780071641555.

Björnar Tessem and Lars Nyre. The Influence of Social Media Use on Willingness to Share Location Information. In Steven Furnell, Costas Lambrinoudakis, and Javier Lopez, editors, *Trust, Privacy, and Security in Digital Business*, pages 161–172. Springer Berlin Heidelberg, 2013.

Mike Thelwall. Privacy and Gender in the Social Web. *Privacy online: Perspectives on Privacy and Self-Disclosure in the Social Web, New York: Springer*, pages 255–269, 2011.

Eran Toch, Justin Cranshaw, Paul Hankes Drielsma, Janice Y Tsai, Patrick Gage Kelley, James Springfield, Lorrie Cranor, Jason Hong, and Norman Sadeh. Empirical models of privacy in location sharing. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, Ubicomp '10, pages 129–138, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-843-8.

Janice Y Tsai, Lorrie Cranor, Serge Egelman, and Alessandro Acquisti. The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, 22(2):254–268, 2010.

Janice Y Tsai, Patrick Kelley, Paul Drielsma, Lorrie Faith Cranor, Jason Hong, and Norman Sadeh. Who's viewed you?: The impact of feedback in a mobile location-sharing application. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2003–2012, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7.

Sherry Turkle. *Life on the Screen: Identity in the Age of the Internet*. Simon & Schuster Trade, 1995. ISBN 0684803534, 9780684833484.

Bibi van den Berg and Ronald Leenes. Audience Segregation in Social Network Sites. *2010 IEEE Second International Conference on Social Computing*, (216483):1111–1116, August 2010.

Sami Vihavainen, Airi Lampinen, Antti Oulasvirta, Suvi Silfverberg, and Asko Lehmuskallio. The clash between privacy and automation in social media. *IEEE Pervasive Computing*, 13(1):56–63, Jan 2014. ISSN 1536-1268.

Jessica Vitak. The Impact of Context Collapse and Privacy on Social Network Site Disclosures. *Journal of Broadcasting & Electronic Media*, 56(4):451–470, October 2012. ISSN 0883-8151.

Nevena Vratonjic, Kevin Huguenin, Vincent Bindshcaedler, and Jean-Pierre Hubaux. How Others Compromise Your Location Privacy: The Case of Shared Public IPs at Hotspots. In *Privacy Enhancing Technologies Symposium (PETS 2013)*, 2013.

James Waldo, Herbert S Lin, and Lynette I Millett. Thinking About Privacy: Chapter 1 of Engaging Privacy and Information Technology in a Digital Age. *Journal of Privacy and Confidentiality*, 2(1):19–50, 2010.

Keqin Wang, Shurong Tong, Lionel Roucoules, and Benoit Eynard. Analysis of data quality and information quality problems in digital manufacturing. In *4th IEEE International Conference on Management of Innovation and Technology*, pages 439–443, Sept 2008.

Richard Y Wang and Diane M Strong. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4):5–33, 1996.

Shaojung Sharon Wang and Michael A. Stefanone. Showing Off? Human Mobility and the Interplay of Traits, Self-Disclosure, and Facebook Check-Ins. *Social Science Computer Review*, pages 1–21, April 2013. ISSN 0894-4393.

Yang Wang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, and Lorrie Faith Cranor. Privacy nudges for social media: An exploratory facebook study. In *Proceedings of the 22Nd International Conference on World Wide Web Companion*, WWW'13 Companion, pages 763–770, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-2038-2.

Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. "i regretted the minute i pressed share": A qualitative study of regrets on facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, SOUPS '11, pages 10:1–10:16, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0911-0.

Samuel Warren and Louis Brandeis. The Right to Privacy. *Harvard Law Review*, 4(5), 1890.

Jen Webb, Tony Schirato, and Geoff Danaher. *Understanding Bourdieu*. SAGE Publications Ltd, 2002. ISBN 0761974636, 9780761974635.

Daniel Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler, and Gerald Jay Sussman. Information accountability. *Communications of the ACM*, 51(6), 2008.

Alan Westin. *Privacy and Freedom*. New York:Atheneum, 1967. ISBN 0370013255, 9780370013251.

Jason Wiese, Patrick Gage Kelley, Lorrie Faith Cranor, Laura Dabbish, Jason I. Hong, and John Zimmerman. Are you close with me? are you nearby?: Investigating social groups, closeness, and willingness to share. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 197–206, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0630-0.

Dmitri Williams. On and Off the 'Net: Scales for Social Capital in an Online Era. *Journal of Computer-Mediated Communication*, 11(2):593–628, January 2006. ISSN 1083-6101.

Craig E Wills and Mihajlo Zeljkovic. A personalized approach to web privacy: awareness, attitudes and actions. *Information Management & Computer Security*, 19(1):53–73, 2011. ISSN 0968-5227.

Heng Xu, Hock-Hai Teo, Bernard C. Y. Tan, and Ritu Agarwal. The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services. *Journal of Management Information Systems*, 26(3):135–174, December 2009. ISSN 0742-1222.

Seounmi Youn and Kimberly Hall. Gender and Online Privacy among Teens: Risk Perception, Privacy Concerns, and Protection Behaviors. *CyberPsychology & Behavior*, 11(6):763–765, 2008.

Alyson Leigh Young and Anabel Quan-Haase. Privacy Protection Strategies on Facebook. *Information, Communication & Society*, 16(4):479–500, May 2013. ISSN 1369-118X.

Aristea M Zafeiropoulou, David Millard, Craig Webber, and Kieron O'Hara. Privacy Implications of Location and Contextual Data on the Social Web. In *ACM Web Science Conference*, pages 1–4, Koblenz. Germany, 2011.

Aristea M Zafeiropoulou, David E Millard, Craig Webber, and Kieron O'Hara. Unpicking the privacy paradox: Can structuration theory help to explain location-based privacy decisions? In *Proceedings of the 5th Annual ACM Web Science Conference*, WebSci '13, pages 463–472, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1889-1.

Aristea M Zafeiropoulou, Kieron O'Hara, David E Millard, and Craig Webber. Location Data and Privacy: A Framework for Analysis. In Bernard Stiegler, editor, *Réseaux sociaux : Culture politique et ingénierie des réseaux sociaux*, pages 185–200. FYP EDITIONS, 2012.

Kathryn Zickuhr. Location-Based Services. Technical report, Pew Research Centre, Washington D.C., 2013a.

Kathryn Zickuhr. Who's not online and why. Technical report, Pew Research Centre, Washington D.C., 2013b.

Kathryn Zickuhr and Aaron Smith. Digital differences. Technical report, Pew Research Centre, Washington D.C., 2012.

J Christopher Zimmer, Riza Ergun Arsal, Mohammad Al-Marzouq, and Varun Grover. Investigating online information disclosure: Effects of information relevance, trust and risk. *Information & Management*, 47(2):115–123, March 2010. ISSN 03787206.

Michael Zimmer. Privacy on Planet Google : Using the Theory of "Contextual Integrity" to Clarify the Privacy Threats of Google's Quest for the Perfect Search Engine. *Journal of Business & Technology Law*, 3(1), 2008.

Jolene Zywica and James Danowski. The Faces of Facebookers: Investigating Social Enhancement and Social Compensation Hypotheses; Predicting Facebook and Offline Popularity from Sociability and Self-Esteem, and Mapping the Meanings of Popularity with Semantic Networks. *Journal of Computer-Mediated Communication*, 14(1):1–34, October 2008. ISSN 10836101.