Is your information sensitive?

creating a new measure for the sensitivity of personal information and predicting information disclosure

Nab, B.

Award date:
2013

Eindhoven, October 2013

Is your information sensitive?

Creating a new measure for the sensitivity of personal information and predicting information disclosure

Bram Nab

Student identity number 0610715

in partial fulfilment of the requirements for the degree of

Master of Science
in Human-Technology Interaction

Department of Industrial Engineering and Innovation Sciences
Eindhoven University of Technology

Supervisors:
dr. ir. A. Haans TU/e, IEIS, Human-Technology Interaction
prof. dr. W. A. IJsselsteijn TU/e, IEIS, Human-Technology Interaction

Keywords
Information Disclosure - Information Privacy - Campbell's Paradigm - Privacy Paradox

The question isn't, 'What do we want to know about people?',
It's, 'What do people want to tell about themselves?'

- Mark Zuckerberg (Founder and CEO of Facebook)

Abstract

Current research about information privacy assumes a causal attitude behavior relationship. From this relationship follows that someone who states to value privacy highly, a high privacy concern, will be protective of his or her information and will thus choose not to disclose personal information to others. In practise however people are consistently found to report to value privacy highly yet still disclose large amounts of information about themselves. This discrepancy between stated attitude and behavior is known as the privacy paradox. In this thesis we abandon the causal relationship between attitude and behavior and instead adopt Campbell's paradigm (Kaiser et al. 2010) which follows an axiomatic relationship between attitude and behavior. We performed two studies to explore the applicability of this paradigm in privacy related research.

In study one we explored whether sensitivity is an invariant attribute of information. For this an online questionnaire was created in which we followed a facetted approach, linking different types of personal information, e.g. your bank balance, with different stakeholders to whom the information would be disclosed, e.g. your municipality. This list of questions was ordered transitively in terms of their likelihood of engagement using Rasch-modelling. Results indicated that we can describe sensitivity of information independent of differences between individuals. Results further indicated an interaction between information type and stakeholder, meaning that we cannot describe sensitivity of information independent of whom the information is disclosed to.

In study two we compared the predictive power of the Rasch-model's estimate for the willingness to disclose information with that of the general concern for information privacy index (Smith, Milberg, & Burke, 1996) and the privacy segmentation index (Kumaraguru & Cranor, 2005). Using a cover story about the evaluation of a health application actual disclosure of information was measured. Our results showed poor predictive power for all three privacy measures. Amongst other explanations for this lack of predictive power, we discuss a ceiling effect in the data for our measure of information disclosure.

This thesis provides a first exploratory step towards the application of Campbell's paradigm in privacy research.

Preface

Privacy, even now at the end of this thesis project I am confused by you. I began with the goal of understanding why I see people act so seemingly inconsistent around me. Why is it that we are outraged when the mailman opens our letters yet are perfectly fine with our email provider scanning through our emails? With over 4 million hits for the term "privacy" on Google Scholar it is clear that privacy is hot. Whilst browsing through this privacy literature it dawned to me that my simple question is not so simple in the context of privacy. Something as basic as a definition of what privacy entails leads to vast discussions between authors and scientific fields.

Throughout this project I have consistently been challenged by the complexity of the subject of privacy. Meanwhile privacy became an even hotter subject, with its peak in early June when Edward Snowden gave the general public insight into the NSA's surveillance methods, PRISM. I strongly belief that research about privacy is now more important than ever; too much information is collected about us with still too little understanding about the consequences, or even what privacy means to people. After little over 7 months of work on this project, my thesis provides a new and in this field unique perspective on privacy and specifically the sensitivity of personal information. I hope research will continue and that one day privacy no longer confuses us.

During my struggles with privacy I received help from my two supervisors, Antal and Wijnand. You have both very much helped me tackle this thesis. Whether it was the much needed explanations of the Rasch-model, a critical question during one of our meetings, or feedback on draft versions. Thank you for your support.

I also wish to specifically thank my friends from "Awesome" and my fellow graduates in the "Awesome office". I have bored you to death with my comments about how difficult privacy is and you have only complained about it occasionally. I still think AH coffee tastes better than Euro Shopper, but we will settle this debate another time. I hope to see many more awesome events in the future.

I finally wish to thank my family for all their help and support throughout this project, without you this project would most likely have lasted forever.

Contents

Chapter 1. Introduction

Researchers have struggled with the concept of privacy for the past decades, yet we are living in a society in which we share more information about ourselves than ever before. Whether we provide fingerprint scans for identity documents, scan our personalized discount card at a local supermarket, or share the photos of yesterday's event through Facebook, everyone is sharing information about themselves with the rest of the world. In today's society, this personal information has become a valuable commodity; one that is being collected by different stakeholders including governments and companies. For most people it remains unclear exactly what information is being stored, by whom, and for what purposes. This often leads to people questioning their privacy. If we wish to provide people with means to better comprehend this collection of information we must first return to the question of what personal information is, and how we can structure it.

The privacy of information is traditionally studied following the causal link between attitude and behavior (Ajzen, 1991). The underlying assumption here is that people's attitude towards privacy can predict their disclosure behavior. In privacy research attitudes are measured through concerns, and privacy behaviour is assessed through surveys, or elicited in experiments. These studies report different valuations of information depending on who is asked to disclose the information (Huberman et al. 2005; Grossklags et al. 2007), to whom the information is disclosed to (Cvrcek et al. 2006), and also what type of information is being disclosed (Hui, Teo, & Lee, 2007).

An important problem in research about information privacy is that the measured concerns do not explain the actual disclosure behavior. It is consistently found that people report privacy as being important to them (Spiekermann, & Cranor, 2009), yet these same people seem not to hesitate when it comes to disclosing their information for very small rewards (e.g. Grossklags et al. 2007). This inconsistency between attitude and behavior is known as the privacy paradox (Norberg, Horne, & Horne, 2007).

In this thesis we will first take a closer look at the privacy paradox and discuss several explanations for this phenomenon. We will then provide an alternative framework for privacy research: Campbell's paradigm (Kaiser, Byrka, & Hartig, 2010). We then report two studies in which we first apply Campbell's paradigm to the question of disclosing information, and, in a second study, validate the results of this application and compare them with currently popular privacy concern measures.

Chapter 2. Theoretical Framework

2.1 Privacy Dimensions

The concept of privacy is being studied in different scientific fields (e.g. Economics, Law, Psychology, Human Computer Interaction). Between these disciplines there is not one clear definition of what privacy entails and how it can be operationalized. Smith, Dinev, and Xu (2011) argue that much of the confusion surrounding the conceptualization of privacy originates from this interest from different fields, and that the concept of privacy has different meanings across disciplines, such as a right of entitlement (in law literature), a state of limited access or isolation (in environmental psychology), and control over information (in information systems literature).

Attempts have been made to describe privacy through a set of dimensions, independent of a specific discipline. A large body of authors have described different sets of dimensions of privacy most popular being Burgoon et al (1989), DeCew (1997) and Clarke (1999). Clarke (1999) and DeCew (1999) both have a focus on personal space and the different traces that are left behind by people, in the form of bits of data that can be linked back to them. Burgoon et al. (1989) have a slightly broader scope by including, for example, a psychological dimension to privacy. This dimension includes the right to determine when and with whom an individual will share their thoughts. Despite this broader scope there still is a large overlap in the different dimensions of each of these authors; all three frameworks include a dimension which relates to how an individual can control and determine how, when, and to what extent information about the self will be released to others. This informational or personal data dimension closely resembles Westin's original definition of privacy as "*An individual's right to control, edit, manage, and delete information about them and decide when, how, and to what extent information is communicated to others*" (Westin, 1967). Central to this dimension is the desire to keep personal information out of the hands of others, to make sure information is stored safely and free of errors, it is a concern for privacy (Buchanan, Paine, Joinson, & Reips, 2007). It is this dimension of information privacy with its concept of controlling how one's personal information is acquired and used (Pavlou, 2011) that is focused on in this study.

Table 1. The dimensions of privacy of Burgoon et al (1989), DeCew (1997) and Clarke (1999).

Burgoon et al. (1989)	**DeCew (1997)**	**Clarke (1999)**
Physical dimension	Informational dimension	Privacy of the personal body
Interactional dimension	Accessibility dimension	Privacy of personal behavior
Psychological dimension	Expressive dimension	Privacy of personal communication
Informational dimension		Privacy of personal data

2.2 The privacy paradox

Existing privacy research follows the attitude behavior paradigm in which someone's attitude can predict and explain variance in actual behavior (e.g. Norberg, Horne, & Horne, 2007). In the context of privacy this would translate into a scenario where someone states to highly value privacy (the attitude) and is reluctant to disclose personal information (the behavior). While this relationship between attitude and behavior at first may seem like an accurate representation of reality practice teaches us differently.

It is consistently found that people report privacy as being important to them (Spiekermann, & Cranor, 2009). IBM (1999) conducted a large multi-national survey (Germany, United States, and the United Kingdom) with a sample of approximately 3000 adults. Almost all their participants (92%) indicated to be concerned about threats to their personal privacy when using the internet. Similarly, a large amount of participants (94%) indicated to be concerned about possible misuse of their personal information. Harris Interactive (2002) provide similar results. In their web-based survey with 1529 participants a majority (79%) agreed that they have lost all control over how their personal information is collected and used by companies, indicating a clear concern for their personal information.

Following the attitude-behavior relationship a strong concern for privacy should lead to privacy protective behavior and a reluctance towards disclosing personal information. This is however not the case. First, individuals are found to be willing to trade personal information in exchange for small rewards; e.g. a survey of office workers found that 71% of them were willing to disclose their computer password in exchange for a chocolate bar (Infosecurity Europe, 2004). Second, individuals are seldom willing to adopt privacy protective technologies or invest in such technologies; e.g. Grossklags and Acquisti (2007) showed how individuals were not willing to spend as little as 25 cents to protect personal information such as their weight.

The attitude-behavior relationship would lead us to expect a strong correlation between expressed atittudes and actual behavior, yet individuals consistently express strong privacy concerns but behave in ways contradictory to these concerns. This phenomenon is known as the privacy paradox (Norberg, Horne, & Horne, 2007). In the following sections different explanations for the existence of the privacy paradox will be discussed covering both the attitude and behavioral side of the relationship as well as contextual and theoretical limitations, before introducing an alternative framework.

2.3 Disclosure Behavior

In privacy research, privacy related behavior is often measured through the value people assign to a piece of personal information. Following this economic perspective of privacy, consumers would be willing to disclose information about themselves in exchange for specific benefits (e.g. money) and similarly choose to keep other types of information to themselves if they do not expect to receive suitable benefits (Pavlou, 2011).

In research this is translated into the popular approach of asking people to disclose information about themselves via auction experiments. Here participants are asked for what price they would be willing to sell personal information such as their weight or income. These experiments have demonstrated that large individual differences exist between people's valuations of their personal information (e.g. Huberman, Adar, & Fine, 2005). There are however equally large differences between studies. Where Huberman et al. (2005) find an average price of €58 to disclose information about age, Grossklags and Acquisti (2007) find that all of their participants were willing to disclose their age for €0.25. While it should be noted that these studies follow a different methodology in terms of how the question was presented to their participants; name an acceptable price for the information (Huberman et al., 2005) vs. would you sell the information for this price (Grossklags and Acquisti, 2007), making a direct comparison between the two studies difficult. It is however clear that the difference in price between the two studies is substantial.

Interestingly Huberman et al. (2005) further demonstrate that participants demand more money for information about their weight when they perceive themselves as heavier than other participants, and that younger participants demand less money for disclosing their age than older participants do. These studies indicate that personal information may not be experienced as something with a fixed and stable price, instead the value of the information is highly dependent on contextual factors. A more detailed discussion of the influence of context can be found in the next section.

The study by Huberman et al. (2005) is an example of a willingness to accept approach, in which the individual states a price for which they are willing to accept to disclose the information. Alternatively, one can also ask an individual to indicate how much they are willing to pay in order to protect their information (i.e., to prevent disclosure). Rose (2005) found that while people may express a high privacy concern they are not automatically willing to pay to protect their information: only 47,5% of the participants expressed to be willing to pay to protect their information, while they did express to be concerned about the privacy of their information. Similarly Grossklags and Acquisti (2007) found that their participants were willing to accept even very small sums of money, as low as 25 cents, to sell information, yet were not willing to pay similar prices to prevent disclosure of the same information.

The described studies illustrate that while people are able to assign an economical value to personal information and adjust their behavior accordingly, this value is far from constant. Acquisti, John, and Lowenstein (2009) further argue that if we are not willing to spend a few cents to protect our privacy yet accept to sell the same data for a similar price a problem occurs in our ideas about the valuation of privacy. Following the economic perspective on privacy means that personal information should have a monetary market value, which individuals are able to define and act upon. We question if individuals are able to do this, to express what value information has to them in a monetary manner. The willingness to accept and willingness to sell methodologies result in different values for the same information, further questioning the possibility to express the value of information in monetary units. If we cannot accurately measure privacy related behavior we cannot begin to describe a relationship between attitude and behavior, let alone speak of a paradox between the two. In the next section will cover the influence context has on the value of personal information. It will further illustrate the problems researchers face when attempting to measure privacy related behavior.

2.4 Contextual Factors

In the process of deciding whether or not to disclose information an individual does not merely consider the value of the information, but the context in which the information is disclosed should also be considered. Context should include: what, when, where, with whom (third parties), and why information is disclosed (Bansal, Zahedi, & Gefen, 2010).

Consumer's beliefs and behavioral responses to privacy threats depend on the type of information requested. A study Phelps, Nowak, and Ferrell (2000) demonstrates that consumers are more willing to disclose demographic and lifestyle information and are less willing to disclose financial information. A survey study by Hui, Teo, and Lee (2007) find similar sensitivity ratings for types of information. However contradicting the results of Phelps et al. (2000) sensitivity had no significant influence on actual disclosure of information. Findings of Carrascal, Riederer, Erramilli, Cherubini, and Oliveira (2011) suggest that we value offline personal information, such as our age, more valuable than online personal information, such as our browsing behavior. Huberman, Adar, and Fine (2005) demonstrate an interaction between type of information and individual characteristics in their study in which people with a high body-mass index (BMI) demanded significantly more payment to disclose information about their BMI than those with a lower BMI value.

Not only the type of information influences the demanded benefit for disclosure but also to whom the information is being disclosed. In a study by Cvrcek, Kumpost, Matyas and Danezis (2006) information about cell phone use was being collected for the period of one month, after which participants were asked how much money would be acceptable as a refund for this data to be used by either a commercial company or for academic purposes. The acceptable refund was twice as high for commercial use demonstrating that the perceived value of personal information is also depending to whom the information is being disclose to.

That context also influences actual behavior was demonstrated by Tsai, Egelman, Cranor, and Acquisti (2011). In their experiment participants were asked to make a purchase online with their own money using one of four websites. The websites differed in privacy levels, as indicated by a five-point scale, as well as the price of the product to be purchased, where higher privacy levels were linked to higher prices. However in only one of the two conditions was this difference in privacy levels revealed to the participant. In the other condition the scale was not explained as an indicator for the level of privacy, instead it was introduced as another non-relevant function, yet no changes were made to the different prices for each website; Thus creating two conditions, one with information about the privacy level of a website and one without such information. A significant difference was found between the two conditions. Participants in the privacy condition bought products at websites with higher privacy levels than participants in the non-relevant condition. Participants were thus willing to pay for higher levels of privacy, once this information was made salient to them. This finding illustrates that contextual information about privacy can also influence purchase behavior.

In sum we have seen that large differences exist between studies, that different methodologies lead to different valuations of information and that types of information are valued differently depending on the specific context in which the behavior occurs. We will now take a more detailed look at what the consequences are of adopting the economic perspective for privacy research. More specifically

we will look at the privacy calculus, the decision process in which an individual weighs the costs and benefits before deciding whether or not to disclose information.

2.5 Privacy Calculus

The privacy calculus follows the economic perspective that we can think of personal information as a commodity that can be sacrificed in return for benefits, as such personal information has become a basic commodity that can be owned and used by others (Joinson & Paine, 2006). It assumes a rational process in which an individual weighs the costs and benefits of sharing a piece of personal information, and, depending on whether the benefits outweigh the costs, decides to disclose the information or not. Research by Phelps, Nowak, and Ferrell (2000) demonstrates this process in practise. They find that when participants were offered shopping benefits (e.g. discounts) in return for disclosing personal information, they considered the benefits greater than the costs and choose to disclose the information.

Benefits are not necessarily of a monetary nature. Social adjustment benefits, for example disclosing information about sexual orientation to establish belonging to a social group, can also have a positive effect on intended disclosure behavior (Lu, Tan, & Hui, as cited in Smith, Dinev, & Xu, 2011). Social media are a clear example of where costs related to disclosing information are negated by the benefits. Facebook users perceive the benefits of using Facebook greater than the observed risks of disclosing personal data (Debatin, Lovejoy, Horn, & Hughes, 2009). We are thus willing to trade information about ourselves in return for benefits, as long as these benefits outweigh the costs of disclosing the information.

In order to calculate the costs of disclosing information an individual has to involve an assessment of the likelihood of the negative consequences as well as the perceived severity of these consequences (Smith, Dinev, & Xu, 2011). We thus need to know what risks are involved in disclosing information. In today's society however, information is gathered as a resource, and the purpose of the data collection can often not be specified until it is used, if used at all, at some unpredictable time in the future (Belloti, 1997). This makes it very difficult to rationally calculate a privacy risk. This is especially true for today's online social networks, which are perhaps best described as online public space (Öqvist, 2009). The online public space has four unique properties (Boyd, 2007): (a) persistence, once published it will exist indefinitely, (b) searchability, unlike the physical world the online public space is easily searchable with tools as Google, (c) replicability, anyone can easily copy things from one place in the public space and publish it at another, (d) invisible audiences, it is unknown who is watching your actions in the public space, or who will in the future. With properties like these it is questionable if we can go through a complete rational process that includes all potential risks and their consequences, and if we are able to compare them to the much easier to comprehend benefits.

Despite the complexity of such decision making process, it is clear that individuals are willing to trade personal information for benefits as long as at the moment of the decision the benefits are perceived as greater than the costs. Within the attitude behavior relationship this means that an individual can value privacy highly, yet still choose to disclose information because of the perceived benefits. Thus what we may first consider as inconsistent behavior may in a more careful analysis be considered as consistent behavior which has been influenced by presented short-term/immediate benefits. We can thus add the privacy calculus, with its assessment of risks and benefits, to the earlier discussed contextual influences that impact behavior. Creating yet another emphasize for the

need to carefully analyze the context in which information is disclosed before coming to the conclusion of whether or not the behavior is inconsistent with the attitude.

Until now we have only focused on the behavior side of the attitude behavior relationship, in the next section we will focus on how attitudes are measured and how this measurement also contributes to the privacy paradox.

2.6 Privacy Concerns

The complex nature of privacy makes it difficult to measure attitudes directly, it is very difficult to answer what your attitude is towards privacy as a whole on a five point scale. Researchers instead rely on the measurement of privacy concerns as a proxy for the general attitude towards privacy (Smith, Dinev, & Xu, 2011). The underlying assumption here is that if someone is concerned about what happens with their personal information they also value privacy highly. Different instruments have been developed to measure privacy concerns, some focusing specifically on concern for information privacy on the internet (e. g. Malhotra, Kim, & Agarwal, 2004) and others focusing on broader levels of information privacy concern.

One of these broader instruments is the general concern for information privacy (GCIP) developed by Smith, Milberg and Burke (1996). The GCIP is currently the most popular instrument to measure information privacy concerns. The instrument consists of fifteen questions divided into four sub-scales of concern for information privacy: collection of personal information, errors in the collected information, secondary use of the information, and unauthorized access of others into your personal information. The instrument suggests that someone with a higher concern for information privacy will perceive that: (a) too much information about them is being collected, (b) much of the collected data contains errors and is therefore inaccurate, (c) organizations use personal information for other purposes than agreed upon, and (d) organizations often fail to provide adequate protection of the personal information stored in their databases.

An alternative popular instrument is the Privacy Segmentation Index (PSI; Kumaraguru & Cranor, 2005). Rather than calculating an individual's privacy concern score like GCIP it will categorize individuals into one of three groups. Privacy fundamentalists, who view privacy as an especially high value that they feel very strongly about; privacy pragmatists, who also have strong feelings about privacy but can also see the benefits from surrendering some privacy in situations and finally privacy unconcerned, those who have no real concerns about privacy or about how other people and organizations are using information about them. Individuals are categorized based on their responses on a short three-item survey.

The two presented instruments (GCIP and PSI) are consistently used throughout the literature to measure an individual's attitude towards privacy and are then used to explain differences in behavior between individuals. Both instruments also consistently find that a high privacy concern does not necessarily lead to privacy protective behavior (i.e., the privacy paradox; Norberg, Horne, & Horne, 2007). This paradox may, however, be an artifact originating from the way concerns are measured in these instruments. Both instruments make use of evaluative statements such as: "*It usually bothers me when companies ask me for personal information".* Behind these statements is the, perhaps implicit, assumption that personal information is one clearly defined concept that has the same

meaning and value for different individuals. But what constitutes personal information: Do your address, your age, or perhaps information about your debts belong to this category? Research has demonstrated that indeed people do not perceive all types of information as equally sensitive (e.g. Phelps, Nowak, & Ferrell, 2000; Hui, Teo, & Lee, 2007) and that the response to such a statement is thus greatly influenced by what type of personal information the individual has in mind at that specific moment.

Secondly, concerns are measured at a global level whereas behavior is measured at a specific level. For example, concerns are measured with the GCIP instrument and the scores are then used to explain variance in whether or not someone is willing to disclose their address when making a purchase (Berendt, Günther, & Spiekermann, 2005). Yet if we wish to explain behavior, we should measure the attitude in relation to the particular behavior in that particular context (Ajzen, 1991). In the above example this is not the case, we compare a global attitude with a specific behavior. This theoretical problem of using general attitudes to explain variance in specific behavior has been recognized by Fishbein and Ajzen (2004) and is known as an "evaluative inconsistency". They argue that as long as we compare global attitudes with specific behavior we cannot expect to find strong correlations. According to Azjen (1991), if we wish to use attitudes to explain behavior we should measure both at a specific level. Thus the small correlations found between an attitude and behavior that are known as the privacy paradox may very well have been caused by this evaluative inconsistency instead of inconsistencies in behavior.

2.7 Returning to the Privacy Paradox

We have introduced the causal attitude behavior relationship that has become common in privacy research. While the relationship would expect strong correlations between an individual's expressed attitude and their actual behavior research repeatedly finds that individuals frequently show behavior that is seemingly inconsistent with their attitude, this we know as the privacy paradox. We have seen that there are several possible explanations for the paradox. For the behavior side of the relationship we saw that studies following auction methodologies lead to large differences in valuations of the same piece of information making it difficult, if not impossible, to define the true value of the piece of information. We further saw the large influence context has on the value of information. Information is valued differently depending on whom it is disclosed to, or what type of information is asked. The privacy calculus introduced the contextual variables of costs and perceived benefits that influence the decision process of whether or not to disclose the information. We suggested that behavior that is considered as inconsistent with an attitude may alternatively be explained by perceived benefits that are associated with disclosure rather than a deviation from the attitude. For the measurement of attitudes we discussed how currently popular instruments have defined the term personal information too broadly, which may lead to different interpretations between individuals. Finally we discussed the implications of global level attitudes to explain specific behavior.

Taken together, we can conclude that the causal attitude behavior relationship is a too simple approximation of reality. If we wish to study privacy and be able to make predictions about the behavior of individuals we will need to adopt an alternative theoretical framework. In the next section we will introduce Campbell's Paradigm (Kaiser, Byrka, & Hartig, 2010) and illustrate how this paradigm could be applied in privacy related research.

2.8 Campbell's Paradigm

In current privacy research a stated attitude is used to explain variance in behavior. Campbell however argued that verbal claims and other behavioral responses towards an attitude object arise from a single "acquired behavioral disposition" (Kaiser, Byrka, & Hartig, 2010). Thus stating that you find privacy important and choosing not to disclose a certain piece of information both arise from the same behavioral disposition, e.g. the disposition to value privacy. The assumption here is that someone's attitude can become apparent through their behaviors. Therefore we no longer speak of a causal relationship between attitude and behavior. Instead the relationship between attitude and behavior is axiomatic, or self-evident.

In Campbell's view both an expressed attitude and the disclosure of information are considered behaviors. Both these verbal claims and the actual behavior have related costs, such as effort, time, money, or potential risk. Some have relatively small costs, other behaviors however may be much more substantial in costs. Campbell's paradigm assumes that these costs are specific to the type of behavior, and that they do not depend solely or even primarily on the individual. This means that the costs for a specific behavior are assumed to be equal for all individuals. If this assumption holds we can compare individuals based on their acquired behavioral disposition. Following the privacy calculus behavior may also have benefits, these are also equal for all individuals, e.g. the risks involved in disclosing your home address in return for discounts is considered equal for all individuals as are the related benefits in the form of discounts.

In Kaiser et al.'s (2010) revised Campbell's paradigm these costs and benefits are translated into a set of behaviors that can be arranged by their probability of engagement: behavior with a higher cost will be less likely to be engaged in than behavior with lower cost. From this follows that people's attitude can become apparent by the behavior they are willing to perform: the more demanding the behavior performed by a person, the more intensely they cherish the goal implied by the attitude, and vice versa. This also allows for the inclusion of contextual influences, The act of disclosing information in different contexts leads to different probabilities of engagement. We would expect that in a low sensitive context the likelihood of disclosing a piece of information is higher than disclosing the same piece of information in a high sensitive context.

The revised Campbell's paradigm has been applied successfully to the field of ecological behavior by (Kaiser, Wölfing, & Fuhrer, 1999). A list of 50 different ecological behaviors was ordered based on their likelihood of engagement, and Byrka (2009) was able to further differentiate between non-environmentalists and environmentalists, as expressed by changing likelihoods of behavior (see Table 2. for a selection). Campbell's paradigm holds that a behavioral disposition must also be observable in an individual's verbal claims. Byrka (2009) confirmed this, showing that both an individual's evaluative statements and their self-reported behavior can be transitively ordered based on their difficulty, and thus likelihood of engagement.

Table 2. Ten Ecological behaviors, ordered by their likelihood of engagement (for a full list see Byrka, 2009)

Behavior	P[1]
1. I bought solar panels to produce energy	0.02
2. I buy milk in returnable bottles	0.07
3. I boycott companies with an unecological background	0.13
4. I read about environmental issues	0.16
5. I drive on freeways at speeds under 100kph	0.30
6. I drive to where I want to start my hikes	0.49
7. I own energy efficient household devices.	0.72
8. After meals, I dispose of leftovers in the toilet.	0.80
9. I put dead batteries in the garbage.	0.87
10. I buy furniture made from tropical woods	0.91

[1]probabilities are presented for an individual with an average environmental attitude

2.9 The Rasch-Model

The key issue in defining an attitude in a Campbellian sense is finding a class of behaviors that can be calibrated as a Rasch scale (Kaiser et al, 2010). This requires the creation of a set of behaviors which can be transitively ordered in terms of difficulty. A set of behaviors is considered transitive if it follows the following reasoning: If behavior A is more demanding than behavior B, and B is more demanding than behavior C, then A must also be more demanding then C. That the disclosure of different types of information follows this reasoning has been demonstrated in several studies (e.g. Phelps, Nowak, & Ferrell, 2000; Hui, Teo, & Lee, 2007)

The application of the Rasch model to the subject of privacy would lead to a mathematical model of the relationship between the act of disclosing a specific piece of information, the individuals willingness to disclose information in general, and the specific costs that are associated with the act of disclosing the information. This results in the following model (for more details, see Bond & Fox, 2007).

$$ln\left(\frac{p(x_{ni}=1)}{1-p(x_{ni}=1)}\right)=\theta_n-\delta_i$$

The Rasch model calculates the natural logarithm for the odds of disclosing a certain piece of information (i) following an additive function of an individual's willingness to disclose information in general (θ_n) and the specific costs that are associated with the act of disclosing that type of information (δ_i). From this follows that the Rasch model is not only descriptive but it is also prescriptive, because of the transitive ordering of the specific disclosure behaviors the Rasch model prescribes which types of information would be disclosed by an individual. For example, if a person reports to disclose six out of twenty types of information, the Rasch model prescribes which six should be disclosed, they are the among ones with the lowest costs associated with the act of disclosing that type of information. Similarly the types of information with the highest associated costs, will only be disclosed by individuals with the highest willingness to disclose information.

While Campbell's paradigm has been successfully applied to other fields in the social sciences it is to our knowledge never applied in privacy related research. In the next section we will introduce our

research question with related hypotheses for the application of Campbell's paradigm in privacy research.

2.10 The current study

We have seen that current privacy research follows the causal attitude behavior paradigm and that in this paradigm an inconsistency between attitude and behavior is frequently found: the privacy paradox. An argument has been made for Campbell's paradigm as an alternative theoretical framework to study privacy. In order to apply this paradigm we should return to the essence of privacy concerns—the sensitivity of personal information,—and ask ourselves the question what kind of attribute sensitivity is? Is it an attribute of a particular kind of information? If this is the case then we should be able to measure the sensitivity of a piece of information independent of the individual's privacy concern, and independent of contextual factors (e.g., who is collecting the information). In the current study we will address this issue and aim to answer the following research question:

Research Question: *Is sensitivity an invariant attribute of personal information?*

This question will be answered in two studies. In Study one we will test if Campbell's paradigm can be applied in privacy related research. For this we will create a class of behaviors and test if they can be calibrated as a Rasch scale (Kaiser et al, 2010); Thus whether we can create a set of behaviors that can be transitively ordered in terms of difficulty including an estimate for an individual's willingness to disclose personal information, as an alternative measure for privacy concern. In Study two we will look at the predictive validity of this estimate and compare its predictive performance to those of other privacy concern instruments. For this we will create an experiment in which actual disclosure behavior is measured. We will then compare the predictive power of the willingness to disclose personal information estimate with other privacy concern measures. For this we will calculate correlations between the amount of disclosed information and each privacy measure. These two studies combined will provide insight in the applicability and added value of adopting Campbell's paradigm in privacy related research.

Chapter 3. Study One; Measuring information sensitivity

3.1 Research Goal

In this first study we will adopt Campbell's Paradigm to explore the concept of the sensitivity of personal information. We will try to find a set of behaviors that can be calibrated as a Rasch scale, following Campbell's paradigm we thus aim at creating one scale on which both expressed concerns for individual types of information and actual disclosure behavior can be mapped. This leads to the following hypothesis, in which we test if Campbell's paradigm can be applied to this type of data:

***H_1**: Data collected about the disclosure of information fits a Rasch model; it can be mapped on a unidimensional scale in which privacy sensitivity of information and individual concern are invariant (i.e., independent).*

Throughout our discussion of the privacy paradox the influence of contextual variables has frequently been mentioned. If sensitivity truly is an invariant attribute of personal information the inclusion of such a contextual influence should affect the probability of a behaviours in an additive and not an interactive way. In other words, if we compare the disclose of information A and B in context X and Y we would expect that the disclosure of information A would be more difficult than B in both X and Y, and that likewise disclosure of information in context X would be more difficult than information disclosure in context Y, for all information types. This leads to a second hypothesis which assumes a successful test of H_1.

***H_2**: Data collected about the disclosure of information fits a Rasch model; it can be mapped on a unidimensional and additive (i.e., non-interactive) scale in which privacy sensitive information, individual concern, and contextual factors (e.g., the agency or company collecting the data) are invariant (i.e., independent).*

Within our first study we will include two types of contextual influences. First we will look at whom the information is disclosed to, we have already seen how previous research shows an effect for this type of context (e.g. Cvrcek, Kumpost, Matyas, & Danezis, 2006). The second type of context we will study the difference between stating that you accept that information is being collected vs. actual disclosure of information. Byrka (2009) illustrated that in it is generally easier to state that you would do something (evaluative statements) than actually performing the action (self-reported behavior).

For the test of our second hypothesis we wish to split the costs associated with the act of disclosing a piece of information into to the costs that: (1) are associated with the information itself, (2) the stakeholder with whom the information is shared, and (3) the format in which question is presented. For this we need to adapt the Rasch model and create a so called many facet Rasch model instead (Linacre, 2002). This leads to the following model:

$$\boldsymbol{ln}\left(\frac{\boldsymbol{p}(\boldsymbol{x_{nisq}}=\mathbf{1})}{\mathbf{1}-\boldsymbol{p}(\boldsymbol{x_{nisq}}=\mathbf{1})}\right)=\boldsymbol{\theta_n}-(\lambda_i+\lambda_s+\lambda_q)$$

This model calculates the natural logarithm for the odds of disclosing a certain piece of information(***i***), which again follows an additive function of an individual's willingness to disclose information in general ($\boldsymbol{\theta_n}$), the specific costs that are associated with the act of disclosing that type

of information ($\boldsymbol{\lambda_i}$) ,the constrains imposed by disclosing information to a specific stakeholder ($\boldsymbol{\lambda_s}$), and finaly the type of question, either behavioral statement or self-reported behavior, that is being answered ($\boldsymbol{\lambda_q}$).

3.2 Method

3.2.1 Design and Procedure

In this first study an online questionnaire was created and hosted for a period of two weeks. The questionnaire consisted of three sections; (1) evaluative statements about the collection of information, (2) self-reported behavior questions for disclosure of information and (3) currently popular privacy concern instruments. Examined variables were the specific type of information that was disclosed, stakeholders to whom the information was disclosed, and the type of question that was answered (evaluative vs. self reported behavior). Question order was randomized between subjects within each section.

3.2.2 Participants

A total of 316 participants (152 male, 164 female) were recruited. The mean age of the participants was 23.2 (SD= 2.9; range 18-30), with missing data for the age of three participants. Participants were invited to take part in an online study about privacy and were recruited via facebook using snowball sampling and also through the JF Schouten participant database. All participants were native Dutch speakers. After full completion of the questionnaire, participants were enrolled in a lottery with a 20% chance to win 10 euro's.

3.2.3 Measures

3.2.3.1 *Information types & Stakeholders*

Because of the exploratory nature of the study the goal was to cover the full spectrum of sensitive information. Here the results of Phelps, Nowak, and Ferrell (2000) as well as Hui, Teo, and Lee (2007) were used as guidance. A total of twelve information types were selected for the questionnaire (see table 3.). A similar approach was followed in the selection of stakeholders. Here a stakeholder for both the private (Employer) and public domain (Municipality), as well as a stakeholder with a commercial (Google), one with a non-commercial focus (Health Insurance), and a stakeholder with an offline orientation (Acquaintances) and one with an online orientation (Facebook) were selected. A final stakeholder (LinkedIn) was included to allow for a direct comparison between self-reported behavior and evaluative statements. We wanted to avoid to ask both an evaluative statement and a self-reported behavior for the same combination of information type and stakeholder, this to prevent participants from falling for the wish to show consistent behavior in their answers. We do however want to compare evaluative statement with self-reported behavior, for this the final stakeholder is used. Only for this stakeholder both evaluative and self-reported behavior questions were included in the questionnaire. For the stakeholders of Facebook and LinkedIn questions are only presented to those participants whom have an account. For Facebook 296 participants (93.7%) indicated to have an account and for LinkedIn 158 participants (50%) indicated to have an account.

Table 3. The information types and stakeholders.

Information Types	Stakeholders
1) Private telephone number	1) Municipality
2) Intelligence quotient (IQ)	2) Employer
3) Date of birth	3) Google
4) Supported charities	4) Health insurance company
5) Fingerprint	5) Acquaintances
6) Bank balance	6) Facebook
7) Medicine use	7) LinkedIn
8) Magazine subscriptions	
9) Holiday photo's	
10) Highest level of education	
11) Supported party at last elections	
12) Internet history of last month	

3.2.3.2 *Evaluative statements*

For the collection of information a set of 52 evaluative statements was created following a facetted approach with five stakeholders and the twelve types of information. Statements were presented in the format of, "I accept that Stakeholder$_X$ knows my Information type$_X$", e.g. "I accept that my municipality knows my private telephone number". All twelve information types were combined with the first four stakeholder (municipality, employer, Google, and health insurance company), the stakeholder LinkedIn was only combined with the first four types of information. Participants were asked to indicate whether they agreed or disagreed with these statements using six response categories, labelled 'disagree', 'slightly disagree', 'neutral', 'slightly agree', 'agree', and 'skip this question'.

3.2.3.3 *Self-reported behavior*

In the self-reported behavior section the same twelve types of information were used and were combined with the two remaining stakeholders: Facebook and acquaintances. The stakeholder LinkedIn was again only combined with the first four types of information. Questions were presented in the format of, "I have shared Information type$_X$ with Stakeholder$_X$", e.g. "I have shared my private telephone number via Facebook". Participants were asked to indicate whether they had or had not done this or choose to skip the question.

A pilot study revealed that some combinations of information type with stakeholders were considered as unrealistic, impossible or simply irrelevant (i.e. sharing your fingerprint with acquaintances, or your internet history of the past month via facebook) . Our goal was to measure whether information was sensitive enough, as measured through whether it is disclosed or not. A problem occurs when participants indicate they have not disclosed the information because it is not applicable and not because they consider the information as too sensitive. To control for this, those combinations that were considered problematic were excluded from the questionnaire resulting in a remaining 20 item self-reported behavior section.

3.2.3.4 *Privacy concern*

We wish to compare our measure of an individual's willingness to disclose information as an estimate for privacy concerns with currently popular privacy concern measures. For this the general concern for information privacy instrument (Smith, Milberg, & Burke, 1996) was included. This instrument consists of 15 statements, for which participants have to indicate whether they agreed or not with using a five-point scale. The average reliability (Cronbach's alpha) for this instrument was a = .895, indicating good internal consistency.

As a second privacy concern measure the privacy segmentation index was included (Kumaraguru & Cranor, 2005). This instrument consists of three 3 statements, for which participants have to indicate whether they agreed or not with using a four-point scale. Participants are then categorized into one of three groups: Privacy fundamentalists, privacy pragmatists, and privacy unconcerned.

For a full list of questions included in the questionnaire see Appendix A.

3.2.3.5 *Analysis*

Responses for the evaluative statements where recoded into a dichotomous format in which 'disagree' and 'slightly disagree' were coded as 'refute' and 'neutral, 'slight agree', and 'agree' were coded as 'accept'. The 'skip this question' responses were coded as missing data. Responses for the self-reported behavior questions were also coded dichotomously, with the 'skip this question' response again as missing data. This type of dichotomization of responses is not expected to affect the meaning of the statements (Linacre, 2009) and would allow for a straightforward interpretation of further analyses.

The general information privacy concern scores were calculated following the guidelines of Smith et al. (1996). Participants were categorized for the privacy segmentation index following the guidelines of Kumaraguru and Cranor (2005).

The Rasch model test was performed using Winsteps software, version 3.75.1 The Many-facet Rasch model test was performed using Facets software, version 3.71.1. Further analysis of an individual's willingness to disclose information was done using IBM SPSS software, version 20.

3.3 Results

In this section we will first present the results of the first Rasch-model including statistics for person and item fit. We will then present the results of the second, facetted-model. Finally we will compare each participant's willingness to disclose information in general with alternative privacy concern measures.

3.3.1 Model 1

The estimates for perceived costs of disclosing information can be found in Table 4. Following the guidelines of Wright and Linacre (1994) all items fit the model acceptably with *MS-values* ≤ 1.20, and four items in the range of 1.20-1.40. Individuals had an average willingness to disclose information of *M* = -.64 logits (*SE* = .39; range -3.41 to 5.19), with a separation reliability of .89. For eighteen (5.7%) of the 316 participants the model did not fit, indicated by a t-value of t ≥ 1.96. The Rasch model empirically explained 57.5% of the variance in the data (16.8% by persons, 40.7% by items), a perfectly fitting model would also explain 57.5% variance, indicating a very good fit for our model. Because the Rasch-model provides estimates for the of the disclosure of information in a probabilistic manner substantial quantization variance is expected, meaning that a perfectly fitting model would never approximate 100% explained variance (see also, Linacre 2003).

A principal component analysis on the standardized residuals was performed to explore the possibility of another unaccounted factor in our data. An additional factor would results in an additional 2.1% explained variance, because of this we concluded that the 72 items tap into one single factor only. To further test this our sample was split in half and for each half a new Rasch-model test was performed, once for persons with an odd identifier and once for persons with an even identifier. These two estimates were highly similar with r = .98 and $p < .001$ (Figure 1), indicating that the perceived costs of disclosing specific types of information is independent of individual differences. In sum our results confirm H_1, our data about the disclosure of information fits a Rasch-model reasonably well and the sensitivity of information and individual characteristics are independent of each other.

Table 4. Perceived costs of disclosing the information (∂) and Standard Error of Estimate (SE), Mean Square Fit Statistic (MS), and the likelihood of agreeing with the statement for an average person (p). Evaluative statements are presented in standard typeface, self-reported behavior statements in *italic* typeface.

	Item	∂	(SE)	MS	p
1	I accept that Google knows my Bank balance	5.48	.75	.77	.00
2	*I have shared my date of birth via LinkedIn*	4.87	1.02	.99	.00
3	I accept that Google knows my fingerprint	4.21	.44	.87	.01
4	I accept that my health insurance knows my internet history of the past month	4.02	.41	.83	.01
5	I accept that my municipality knows my internet history of the past month	4.02	.41	.86	.01
6	I accept that my employer knows my bank balance	3.86	.39	.94	.01
7	*I have shared my IQ via Facebook*	3.84	.42	1.31	.01
8	*I have shared my private telephone number via LinkedIn*	3.73	.60	1.02	.01
9	I accept that my municipality can see my holiday photo's	3.59	.35	.86	.01
10	I accept that my health insurance can see my holiday photo's	3.37	.32	.85	.02
11	I accept that Google knows which medicines I use	3.37	.32	.90	.02
12	*I have shared which political party I support via Facebook*	2.89	.28	1.16	.03
13	I accept that my health insurance knows my bank balance	2.67	.25	1.03	.04
14	I accept that my health insurance knows which party I supported at the last elections	2.61	.24	.82	.04
15	I accept that my municipality knows my bank balance	2.55	.24	1.01	.04
16	I accept that my employer knows my internet history of the past month	2.33	.22	1.17	.05
17	I accept that Google knows which party I supported at the last elections	2.29	.22	.83	.05
18	*I have shared which magazines I am subscribed to via Facebook*	2.24	.24	1.24	.05
19	*I have shared my bank balance with my acquaintances*	2.19	.21	1.28	.06
20	I accept that my health insurance knows which magazines I am subscribed to	1.68	.18	.85	.09
21	I accept that Google knows which charities I support	1.58	.18	.78	.10
22	I accept that my municipality knows which magazines I am subscribed to	1.46	.17	.84	.11
23	I accept that Google can see my Holiday photo's	1.43	.17	.94	.11
24	*I have shared my private telephone number via Facebook*	1.40	.18	1.40	.12
25	I accept that my health insurance knows which charities I support	1.26	.17	.78	.13
26	I accept that Google knows my IQ	1.20	.16	.81	.14

	Item	∂	(SE)	MS	p
27	I accept that Google knows my privacy telephone number	1.15	.16	1.07	.14
28	I accept that my municipality knows which medicines I use	1.10	.16	.92	.15
29	I accept that Google knows which magazines I am subscribed to	1.02	.16	.83	.16
30	I accept that LinkedIn knows which charities I support	.99	.22	.96	.16
31	I accept that my employer knows which party I supported at the last elections	.97	.16	.98	.17
32	I accept that my employer can see my holiday photo's	.93	.15	1.08	.17
33	I accept that my employer knows which magazines I am subscribed to	.83	.15	.94	.19
34	I accept that Google knows my internet history of the past month	.72	.15	1.10	.20
35	I accept that my health insurance knows my IQ	.49	.14	.89	.24
36	I accept that my employer knows my fingerprint	.45	.14	1.02	.25
37	I accept that my municipality knows which party I supported at the last elections	.43	.14	1.10	.26
38	I accept that my municipality knows which charities I support	.41	.14	.92	.26
39	I accept that my municipality knows my IQ	.28	.14	.87	.28
40	I accept that my health insurance knows my fingerprint	.20	.14	1.05	.30
41	*I have shared which medicines I use with my acquaintances*	.06	.15	1.19	.33
42	I accept that Google knows my highest level of education	-.10	.13	.87	.37
43	I accept that my employer knows which charities I support	-.26	.13	.90	.41
44	*I have shared my IQ with my acquaintances*	-.29	.14	1.13	.41
45	I accept that LinkedIn knows my IQ	-.50	.18	.97	.47
46	I accept that Google knows my date of birth	-.55	.13	.96	.48
47	I accept that my health insurance knows my highest level of education	-.59	.13	.96	.49
48	I accept that my employer knows which medicines I use	-.62	.13	1.10	.50
49	I accept that LinkedIn knows my private telephone number	-.85	.18	1.07	.55
50	*I have shared which political party I supported at the last elections with my acquaintances*	-1.31	.13	1.07	.66
51	I accept that my municipality knows my fingerprint	-1.42	.13	1.00	.69
52	*I have shared which charities I support with my acquaintances*	-1.50	.14	1.06	.70
53	I accept that my municipality knows my private telephone number	-1.57	.13	1.04	.72
54	I accept that my employer knows my IQ	-1.79	.14	1.07	.76
55	*I have shared which magazines I am subscribed to with my acquaintances*	-1.92	.15	1.05	.78

	Item	∂	(SE)	MS	p
56	I accept that my municipality knows my highest level of education	-1.97	.14	.89	.79
57	*I have shared my IQ via LinkedIn*	-2.16	.20	1.05	.82
58	I accept that my health insurance knows my private telephone number	-2.26	.15	1.01	.83
59	*I have shared my holiday photo's via Facebook*	-2.32	.15	1.11	.84
60	*I have shared my highest level of education via Facebook*	-2.87	.18	1.03	.90
61	*I have shared my date of birth via Facebook*	-2.96	.18	1.05	.91
62	I accept that LinkedIn knows my date of birth	-3.42	.29	.91	.94
63	I accept that my health insurance knows which medicines I use	-3.49	.21	1.12	.95
64	*I have shared my holiday photo's with my acquaintances*	-3.53	.21	.96	.95
65	I accept that my employer knows my private telephone number	-3.96	.25	1.02	.97
66	*I have shared my highest level of education with my acquaintances*	-4.98	.39	1.00	.99
67	*I have shared my private telephone number with my acquaintances*	-5.14	.42	.99	.99
68	*I have shared my date of birth with my acquaintances*	-5.83	.58	1.03	.99
69	I accept that my employer knows my date of birth	-5.85	.58	1.02	.99
70	I accept that my health insurance knows my date of birth	-6.26	.71	1.01	1.00
71	I accept that my employer knows my highest level of education	-6.96	1.00	1.00	1.00
72	I accept that my municipality knows my date of birth	-6.96	1.00	.99	1.00

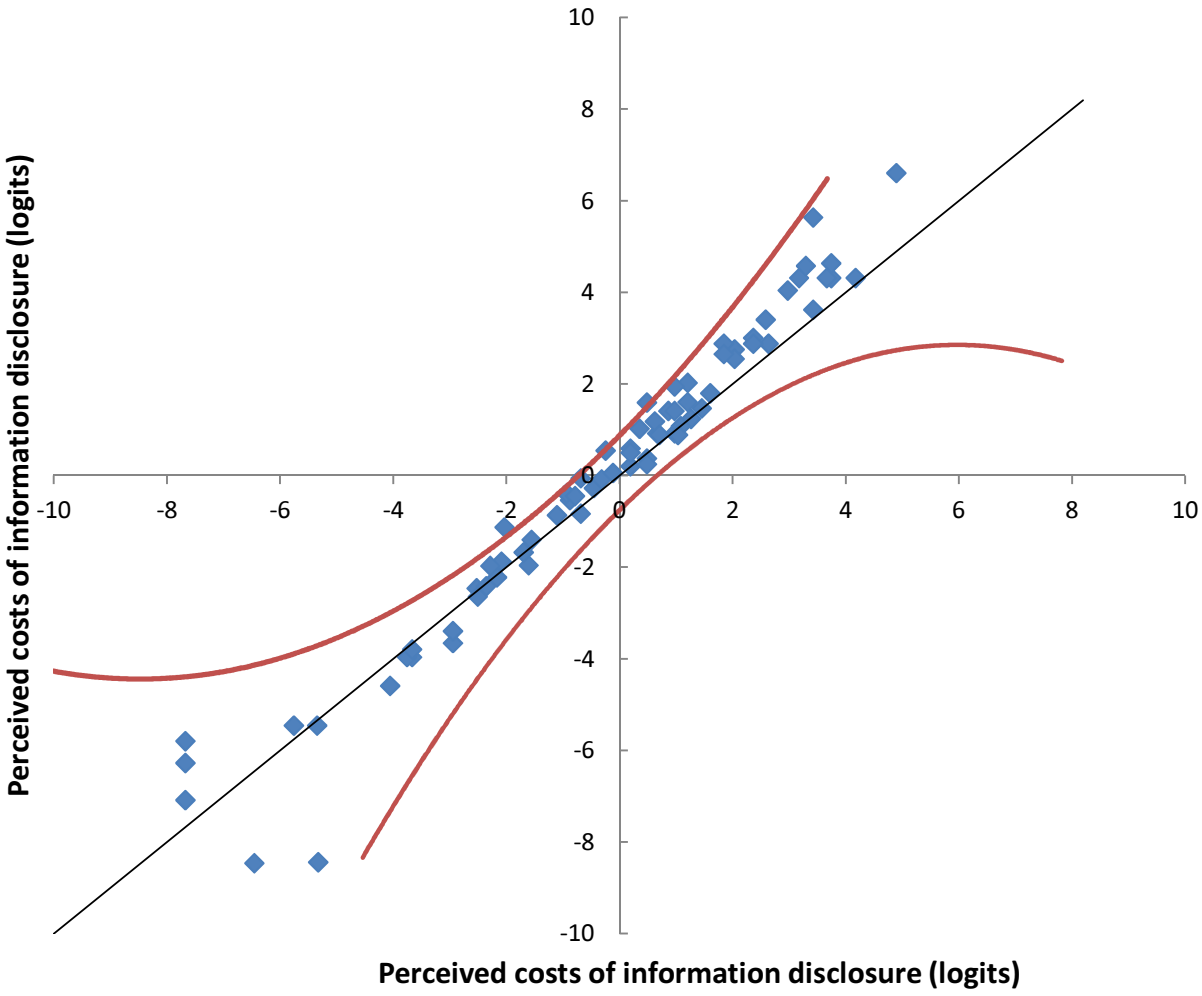

Figure 1. The perceived costs of information disclosure compared for the Odd and Even sample.

The perceived costs of information disclosure, and 95% confidence intervals, for the 72 reported behaviors, estimated for the participants with even identifiers as transformed in the metric of the estimates for the odd participants.

3.3.2 Model 2

The estimates for the specific costs that are associated with the act of disclosing a type of information, the constrains imposed by disclosing information to a specific stakeholder, and estimates for the constrains imposed by the type of question are reported in Table 5a,b, and c. For information type we see that all but one item fit the model well (MS < 1.20). The fingerprint type of information still has an acceptable MS value (MS = 1.31). For the constrains imposed by stakeholder we see that all but one item fit the model well, here the LinkedIn stakeholder has poor fit (MS = 1.43). Both types of questions fit out model well as indicated by an MS value below 1.20. Individual's had an average willingness to disclose information of *M* = -1.22 logits (*SE* = .32; range -2.94 to 3.82). For fourteen of the 316 (4.4%) the model did not fit, again indicated by significant t-values of t > 1.96. A chi-square test revealed a significant interaction between stakeholder and type of information that was disclosed: χ^2 (69) = 3383.5, p<.001. Closer examination of different combinations of stakeholder and type of information further revealed large effects, e.g. the estimated difference between disclosing your holiday photo's to your health insurance and disclosing you medicine use to your health insurance differed 5.82 logits which was significant as shown in a Welch's *t*-test, *t*(604) = 16.44, *p* < .001. Similarly the estimated difference between disclosing your internet history to your employer or to Google differed 4.56 logits which was significant as shown in a Welch's *t*-test, *t*(603) = 14.01, *p* <.001. We thus need to reject our second hypothesis, while our data does fit a facetted Rasch-model reasonably well, the interaction tests reveal a strong interaction effect between type of information and stakeholder, thus contextual effects are not invariant and should be taken into account when we look at the disclosure of information. For our further analysis this means that we will adopt the results of the first model, in the next section we will compare the performance of this first model's estimates of individuals willingness to disclose information with currently popular privacy concern instrument.

Table 5a-c. Perceived costs of disclosing the information/stakeholder/type of question (λ) and Standard Error of Estimate (SE), Mean Square Fit Statistic (MS), and the likelihood of agreeing with the statement for an average person (p)

	Information Item	λ	(SE)	MS	p
1	Bank Balance	3.09	.75	.77	.01
2	Internet history of past month	1.75	1.02	.99	.05
3	Supported party at last elections	.91	.44	.87	.12
4	Magazine subscriptions	.66	.41	.83	.14
5	Supported charities	.31	.41	.86	.19
6	Holiday photo's	.06	.39	.94	.24
7	Fingerprint	-.09	.42	1.31	.26
8	Medicine Use	-.15	.60	1.02	.27
9	IQ	-.20	.35	.86	.28
10	Private telephone number	-1.32	.32	.85	.55
11	Highest level of education	-2.10	.32	.90	.73
12	Date of birth	-2.94	.28	1.16	.86

	Stakeholder	λ	(SE)	MS	p
1	Google	2.09	.05	1.13	.04
2	Health Insurance	.66	.04	.94	.14
3	LinkedIn	.60	.08	1.43	.15
4	Municipality	.59	.04	.88	.15
5	Employer	-.03	.04	.79	.25
6	Facebook	-.86	.06	1.15	.44
7	Acquaintances	-3.04	.05	1.03	.87

	Type of question	λ	(SE)	MS	p
1	Self-Reported Behavior	.92	.04	1.18	.12
2	Evaluative	-.92	.02	.92	.45

3.3.3 Privacy Concern

Individuals reported a general concern for information privacy (GCIP, Smith et al., 1996) of *M* = 5.7 (*SE* = .67; range 3.81 to 7) indicating an above average concern for information privacy. The privacy segmentation index (Kumaraguru & Cranor, 2005) categorized 32 (10%) individuals as privacy fundamentalists, 205 (65%) as privacy pragmatists and 79 (25%) as privacy unconcerned. Our sample consists of more privacy pragmatists and less privacy unconcerned individuals than the population averages reported by Kumaraguru and Cranor (2005), again indicating an above average concern. Scores were standardized for both the GCIP scores and the Rasch-model's estimate of an individual's willingness to disclose information, five outliers were selected as indicated by Z-values of Z > 4, leaving a total of 311 participants in our sample. Normality tests revealed deviations from normality for the estimates of an individual's willingness to disclose information with D(311) = .074, P <.001.

A small correlation was found between the GCIP scores and the estimates for an individual's willingness to disclose information with *rho* = -.276, p < .001. The correlation is negative because the estimates follow an inverse scoring compared to the GCIP, a higher GCIP score indicates a stronger concern for privacy whereas a lower estimate indicates a lower willingness to disclose information. An univariate analysis of variance (ANOVA) with the estimate of the willingness to disclose information as dependent variable, and the privacy segmentation index as factor. No significant effect was found with F(2,308) = .750, *P* = .473 and η_p^2 = .005. A second univariate analysis of variance (ANOVA) with the general concern for information privacy score (GCIP) as dependent variable, and the privacy segmentation index as factor. Here an effect was found with F(2,303) = 5.467, *P* < .01 and η_p^2 = .035. Pair-wise comparisons (LSD) revealed that participants that were categorized as privacy fundamentalists had higher GCIP scores than both pragmatists (*p* < .01) and unconcerned (*p* <.05) individuals, and that pragmatists scored higher than unconcerned individuals (*p* < .01).

To explore the predictive power of Campbell's paradigm we wish to use the estimate for the willingness to disclose information as a predictor for behavior that is not included in the creation of the scale. With the current estimate this is not possible as all behavior was included in the current measure. Therefore a second Rasch analysis is performed using only half of the questions of the questionnaire. The selection of questions for this second analysis was based on the transitive order of questions of the first analysis, we selected all odd-ranked questions from Table 4 (Thus the most difficult, third most difficult, fifth most difficult, etcetera.). In this second analysis individuals had an average willingness to disclose information of *M* = -.75 logits (*SE = .54*; range -4.43 to 6.10), with a separation reliability of .81. This second estimate was highly similar to the first as indicated by a correlation of *rho* = .943, p < .001.

The predictive validity of this second estimate for the willingness to disclose information, the GCIP-scores, the privacy segmentation index and the self-reported behavior questions that were not included in the second estimate are reported in Table 6. For the correlations between the privacy segmentation index and the self-report behavior questions a univariate ANOVA was first performed for each question. The related correlation was then calculated by taking the square root for each of the partial eta squared values. The estimate for the willingness to disclose information had an average correlation of *M* = .21 (*SD* = .13; range 0 to .43), average correlation for the GCIP scores was *M* =-.06 (*SD* = .06; range -.14 to .04), and the average correlation for the privacy segmentation index

was *M* =.07 (*SD* = .03; range .03 to .11). The predictive validity for two specific reported behavior questions and related evaluative statements for the stakeholder of LinkedIn, the estimate for the willingness to disclose information, the GCIP-score, and the privacy segmentation index are reported in Table 7.

Table 6. Predictive validity of self-reported disclosure behavior for the second Rasch-model's estimate for the willingness to disclose information, the general concern for information privacy measure (GCIP), and the privacy segmentation index.

	Willingness to disclose information	GCIP	Privacy Segmentation Index[1]
Item - Stakeholder	**rho**	**rho**	**r**
Private telephone number - Facebook	.128*	-.055	.114
Magazine subscriptions - Facebook	.171**	-.138*	.077
Holiday photo's - Facebook	.338**	-.041	.094
Highest level of education - Facebook	.202**	-.188**	.031
Supported party at last elections - Facebook	.249**	-.044	.031
Supported charities - Acquaintances	.360**	-.082	.089
IQ - Acquaintances	.427**	-.103	.063
Date of birth - Acquaintances	.001	.044	.094
Highest level of education - Acquaintances	.095	-.034	.109
Supported party last elections - Acquaintances	.319**	-.031	.044
Private telephone number - LinkedIn	.127	-.085	.044
Date of birth - LinkedIn	.130	-.017	.070

[1]r values are calculated by taking the square root of the partial eta squared estimates
* p < .05 ** P < .01

Table 7. Predictive validity of self-reported disclosure behavior for the related evaluative statement, the second Rasch-model's estimate for the willingness to disclose information, the general concern for information privacy measure (GCIP), and the privacy segmentation index.

	Specific evaluative statement	Willingness to disclose information	GCIP	Privacy Segmentation Index[1]
item	**rho**	**rho**	**rho**	**r**
Private telephone number	.476*	.127	-.085	.044
Date of birth	.169**	.130	-.017	.070

[1]r values are calculated by taking the square root of the partial eta squared estimates
* p < .05 ** P < .01

3.4 Discussion

In this first study we tested whether we could adopt Campbell's paradigm to explore the concept of the sensitivity of personal information. Our results show that we have been able to create a set of behaviors that can be ordered transitively in terms of difficulty via a Rasch-model. These results confirm our first hypothesis, the transitive order of different behaviors is invariant of individual differences in concern. Providing first support for the application of Campbell's paradigm in the field of privacy research. The large influence of context on the value of information that is consistently found in existing privacy research (e.g. Cvrcek, Kumpost, Matyas, & Danezis, 2006) also became evident in this first study. We found differences between stakeholders as large as 5 logits, which is equal to a difference in likelihood of disclosure of over 80 percent. Our results further show an overall significant interaction effect between type of information and stakeholder to whom the information was disclosed to. This meant a rejection of our second hypothesis, information sensitivity and contextual factors are not invariant. This is an important finding for it means that we should describe information disclosure as the act of disclosing a specific type of information to a specific stakeholder. An individual's privacy concern, as expressed through their willingness to disclose information in general as measured using the Rasch-model showed a small correlation with the general concern for information privacy instrument (GCIP). When looking at the predictive power of the privacy concern instruments for specific self-reported behavior we saw an acceptable predictive power for the estimate for the willingness to disclose information and poor power for the GCIP and privacy segmentation instruments.

While the results are promising and certainly suggest a viable application of Campbell's paradigm in privacy research more validation is needed. The predictive power of the estimate for the willingness to disclose information had to be tested making use of only half questions of our questionnaire, while strongly correlated to the estimate's of the full Rasch-model's estimate, using only half of questions meant a lower separation reliability between individuals. A second limitation in the current study lies in our use of self-reported behavior questions, they are self-reported, they do not measure actual disclosure behavior. The true privacy paradox lies between privacy concern and actual disclosure of information. Real validation of our estimate for the willingness to disclose information should thus come from an independent set of actual behaviors. In our second study we will create such a set of actual behaviors and compare the performance of the complete Rasch-model's estimate with the performance of the GCIP instrument and the privacy segmentaion index in explaining variation in the measured behavior.

A more in-depth discussion of this first study can be found in the general discussion.

Chapter 4. Study Two; Predicting information disclosure

4.1 Research Goal

We concluded our first study with the need for more research to validate the use of Campbell's paradigm in privacy research. Particularly research that looks into actual disclosure behavior was necessary in order to judge the validity of the Rasch-model's estimate for the willingness to disclose information. In our second study we will provide a first step towards such research. In this study we will compare the Rasch-model's estimate for an individual's willingness to disclose information with the general concern for information privacy (GCIP) and the privacy segmentation index. We will compare their predictive power when it comes to the actual disclosure of personal information. We thus need to create an experimental situation in which we can measure the disclosure of personal situation in a natural setting, e.g. participants should not be aware of the fact that we are interested in the disclosure of information. In previous privacy research cover stories have successfully been used to measure disclosure behavior (e.g. Berendt, Günther, & Spiekermann, 2005; Tsai, Egelman, Cranor, & Acquisti, 2011).

In this second study we will also make use of a cover story, we will invite participants to evaluate a currently being developed health application. The application will ask questions related to the current health situation of the participant and based on the answers of the participant provide feedback. This will allow us to measure the information disclosure in a way in which the participants are blind to the real goal of our study.

Our goal of this second study is to compare the predictive power of the three privacy concern measures; the Rasch model's estimate , the GCIP instrument, and the privacy segmentation index. As a measure for information disclosure we will use the amount of answered and unanswered questions, following the assumption that when a question remains unanswered the participant considers that question as too sensitive. From this follows a first hypothesis:

H_3: *The estimate for the willingness to disclose information will better predict the amount of unanswered questions than the general concern for information instrument (GCIP), and the privacy segmentation index, as indicated by a stronger correlation.*

Answering a health related question does not necessarily mean you have disclosed sensitive information. There are types of questions for which only one out of several answer options constitutes the disclosure of sensitive information. For example, is it not considered sensitive to state that you do not have a STD. However answering that you do have a STD is sensitive. This means that merely measuring whether a question was answered or not does not capture the full spectrum for the disclosure of information. We are interested in information disclosure, particularly the disclosure of sensitive information. This leads to a second measure, in which we measure the amount of actual sensitive information that is disclosed. Here we wish to compare the predictive performance of the two measure, which leads to the following hypothesis:

H_4: *The estimate for the willingness to disclose information will better predict the amount of sensitive information that is being disclosed than the general concern for information instrument (GCIP,) and the privacy segmentation index, as indicated by a stronger correlation.*

4.2 Method

In the following section we will first explain the followed methodology in more detail, before we will report and discuss the results of this second study.

4.2.1 Design

In this second study we wish to measure actual disclosure of information. For this a cover story about the evaluation of an online health application was created. Participants were asked to evaluate this application. The health application was set up as an online questionnaire in which participants would answer a set of questions about a specific health area before receiving health advice in the form of simple tips to improve their health situation. The examined variables in this study were the amount of unanswered questions and the amount of sensitive information that was disclosed during the evaluation.

4.2.2 Participants

Participants were sampled from the participants that had completed the full questionnaire of study one. A total of 64 participants (32 male, 31 female, 1 undisclosed) took part in study two. The mean age of participants was 22.7 (*SD* = 2.7; range 18 to 29), with one participant choosing not to disclose their age. All participants were native Dutch speakers. After full completion of the questionnaire participants were enrolled in a lottery with a 20% chance to win 10 euro's.

4.2.3 Apparatus and Procedure

Participants were contacted by email and were invited to participate in a study about a new health application. They were asked to evaluate this new health application called Total Fit. Participants were told the application was currently in development at the Eindhoven university of technology and that the researchers were looking for people willing to test the system. The email concluded with a link to the application. The application itself was created using Limesurvey version 1.90.

At the start of the questionnaire participants were told that application would generate a profile about their current mental and physical health based on the answers that they would provide to the questions. Through this profile they would receive specific feedback based on their own unique situation. Participants were asked to use the application once and were told the researchers were interested in their feedback on this application. Finally participants were assured that their answers would only be used for the purpose of the study and would never be shared with third parties.

The health application consisted of five parts: general information, dietary questions, alcohol and other drug use, sexual activity, and finally questions related to self image. After each part feedback was provide, while participants were told this feedback would depend on their answers all participants received identical tips and feedback. At the end of the application participants were asked to provide feedback about their use of the application via a short questionnaire. After the questionnaire was submitted participants were led to a debriefing web page in which the real purpose of the study was disclosed. It was emphasized that the application is not a real product and was designed only for the purpose of this study. They were told that the provided health tips were generic and were independent of their answers, finally they were given contact information of the experiment leader should they wish to receive more information about the study.

4.2.4 Measures

The questionnaire consisted of 37 questions divided over five categories; 7 general, 5 dietary, 6 alcohol and other drug related, 10 sexual activity, and 9 questions related to self image (A complete list of questions is provided in appendix B). Depending on the information that was asked response options where either open-ended, provided in a yes - no format or presented as a statement with an agree or disagree answer-option. All closed questions had a "skip this question" response option, for the open-ended questions it was emphasized that it was not mandatory to answer each question. For our measure of unanswered questions the "skip this question" responses for the closed questions was coded as unanswered, all other options were coded as answered. For the open-ended questions all blank and incomplete (e.g. some participants only stated only the first letter of their name instead of their full name) responses were coded as unanswered and all complete responses were coded as answered.. 23 of the questions were identified as having sensitive response options, those response options that were identified as sensitive were coded as disclosed information. The remaining responses, including the "skip this question" responses, were coded as not disclosed. For the "*what is your name?*" questions alias answers (e.g. one participant stated his name was Gerard Depardieu) were also coded as not disclosed, as these answers do not actually disclose information about the person.

Three questions were included as control measures. These questions: *"I did not find it difficult to answer the questions", "I considered the questions odd", "I felt comfortable answering the questions"* were added in the evaluation section of the application. Participants had to indicate whether they agreed or not with these statements using a five-point scale.

The estimate for the willingness to disclose information from study one was used with an average of M = -.67 (SD = 1.1; range -2.4 to 1.9) as well as the general concern for information privacy scores with an average of M = 5.6 (SD = .6; range 3.92 to 6.9). Finally the privacy segmentation index's categorization was adopted with 8 out of 64 (13%) privacy fundamentalists, 41 (64%) pragmatists and 15 (23%) privacy unconcerned. All three measures were highly similar to the reported measures of study one.

4.3 Results

4.3.1 Unanswered Questions

Participants had an average amount of unanswered questions of *M* = 1.3 (*SD* = 2.6; range 0 to 15). Correlations between the estimate for the willingness to disclose information, the GCIP-scores, the privacy segmentation index and each of the questions are reported in Table 8 . For the correlations between the privacy segmentation index and the unanswered questions a univariate analysis of variance (ANOVA) was first performed for each question, with the question as dependent variable and the privacy segmentation index as factor. The related correlation was then calculated by taking the square root for each of the partial eta squared values. No significant correlations were found between any of the measures and specific questions. Average correlations were calculated for each of the measures. The willingness to disclose information had an average correlation of *M* = .07 (*SD* = .10; range -.12 to .22), the general concern for information privacy index had an average correlation of *M* = -.01 (*SD* = .09; range -.13 to .15), and the privacy segmentation index had an average correlation of *M* = .16 (*SD* = .06; range .09 to .29). The amount of unanswered questions was summated for each participant and correlated with the three control questions. Three significant correlations were found, participants that indicated they did not find it difficult to answer the questions had less unanswered questions (r = -.262, p = .036), participants that considered the questions odd had more unanswered questions (r = .373, p = .002) and finally participants that felt comfortable answering the questions had less unanswered questions (r = -.357, p =.004). None of these control questions correlated with any of the privacy measures, correlations are reported in Table 9.

Table 8. Predictive validity between unanswered questions and the different privacy measures

item		Willingness to disclose information r	GCIP r	Privacy Segmentation Index[1] r
1.	What is your name	.046	-.014	.286
2.	What is your age	-.078	.081	.134
3.	What is your gender	-.009	-.018	.228
4.	To which ethnic group do you belong	.203	-.133	.294
5.	Do you have a religion	-.027	.089	.118
6.	Do you currently use any medicine	.136	-.134	.094
7.	Are there genetic diseases in your family	.136	-.134	.094
8.	I sometimes feel guilty after eating something unhealthy	.101	.086	.094
9.	On how many days of the week do you eat fish and/or meat produce	.175	-.134	.228
10.	I have once drunk so much that I could not remember everything the day after	-.017	.048	.167
11.	I sometimes drink alcohol because it fits in the social situation and not because I want to	.169	-.034	.134
12.	I have done something when I was drunk that embarrasses me	-.020	-.029	.167
13.	Do you use soft drugs	.030	.154	.122
14.	Do you use hard drugs	.030	.154	.122
15.	Are you currently in a relationship	.136	-.134	.094
16.	Have you ever cheated on someone	-.079	.146	.192
17.	What is your sexual orientation	.091	-.108	.122
18.	Are you currently sexually active	.086	.035	.089
19.	With how many individuals have you had sex in the past year	.117	.092	.1
20.	Have you ever had a sexually transmitted disease	.136	-.009	.228
21.	I have had sex with more than one person within the same day	.208	-.013	.089
22.	I sometimes have unprotected sex	.219	-.034	.089
23.	In case of unwanted pregnancy an abortion is a serious consideration for me	-.043	-.085	.181
24.	Do you currently experience symptoms of incontinence	.194	-.102	.122
25.	I often feel lonely	-.112	-.078	.228
26.	I hate my life	-.009	-.018	.228
27.	I sometimes consider committing suicide	-.012	-.006	.230
28.	I feel confident about my body	.036	.055	.122

[1]r values are calculated by taking the square root of the partial eta squared estimates
* p < .05 ** P < .01

Table 9 Predictive validity between control questions, the amount of unanswered questions, the amount of disclosed information and the different privacy measures.

	Amount of unanswered questions	Amount of disclosed information	Willingness to disclose information	GCIP	Privacy Segmentation Index[1]
	r	r	r	r	r
I did not find it difficult to answer the questions	-.262*	-.004	.014	-.130	.268
I considered the questions odd	.373**	-.240	.015	-.029	.257
I felt comfortable answering the questions	-.357**	.045	.155	-.002	.219

[1]r values are calculated by taking the square root of the partial eta squared estimates
* p < .05 ** p < .01

4.3.2 Disclosed Information

Participants had an average amount of disclosed information of *M* = 6.98 (*SD* = 2.8; range 2 to 14). Correlations between each individual question and the different privacy measures were calculated, correlations are reported in Table 10. No significant correlations were found for the estimate for the willingness to disclose information. The average correlation for this willingness to disclose information was *M* = -.01 (*SD* = .08; range -.16 to .17). For the general concern for information privacy instrument one significant correlation was found with the "*Are you under- or overweight?"* (r = -.304, p = . 015), indicating that participants with a higher concern for privacy were less likely to disclose information for this particular question. The average correlation for the GCIP with all of the questions was *M* = -.06 (*SD* = .10; range -.30 to .10).

For the privacy segmentation index three significant effects were found using univariatie ANOVA with each of the questions as dependent variable and the privacy segmentation index as factor. The privacy segmentation index was found to have a significant for *"I have done something when I was drunk that embarrasses me"* with $F(2,61) = 3.186$, $p = .048$, $\eta_p^2 = .095$. Pair-wise comparisons (LSD) revealed that privacy unconcerned individuals had a significantly higher amount of disclosed information than privacy pragmatists ($p < .05$), no difference was found between privacy unconcerned and privacy fundamentalists individuals, or between fundamentalists and pragmatists. The privacy segmentation index was also found to have a significant effect for *"I have had sex with more than one person within the same day*" with $F(2,61) = 3.383$, $p = .040$, $\eta_p^2 = .100$. Here pair-wise comparisons (LSD) revealed only an effect between privacy pragmatists and unconcerned ($p < .05$), privacy unconcerned individuals being more likely to disclose information for this question. Finally the privacy segmentation index had a significant effect for "*Do you currently experience symptoms of incontinence*" $F(2,61) = 3.813$, $p = .028$, $\eta_p^2 = .111$. Pair-wise comparisons (LSD) revealed that for this questions privacy fundamentalists were more likely to disclose information than the other two groups (both $p < .05$). The average correlation for the privacy segmentation index with all of the questions was *M* = .04 (*SD* = .04; range .001 to .11)

The amount of disclosed information was summated for each participant and correlated with the three control questions. No significant correlations were found. All correlations between the control questions and the different privacy measures, as well as the amount of disclosed information are reported in Table 9.

Table 10. Correlations between disclosed information in each question and the different privacy measures

item		Willingness to disclose information r	GCIP r	Privacy Segmentation Index[1] r
1.	What is your name	-.112	-.179	.028
2.	Do you currently use any medicine	-.036	-.072	.009
3.	Are there genetic diseases in your family	.051	-.044	.027
4.	I am an emotional eater	.037	-.048	.018
5.	I sometimes feel guilty after eating something unhealthy	-.155	-.029	.090
6.	I have once drunk so much that I could not remember everything the day after	-.017	-.242	.066
7.	I sometimes drink alcohol because it fits in the social situation and not because I want to	-.021	.026	.020
8.	I have done something when I was drunk that embarrasses me	-.045	-.164	.095*
9.	Do you use soft drugs	-.055	.101	.058
10.	Do you use hard drugs	-.122	-.071	.021
11.	Have you ever cheated on someone	.023	.040	.065
12.	Have you ever had a sexually transmitted disease	.030	-.008	.009
13.	I have had sex with more than one person within the same day	-.082	.063	.100*
14.	I sometimes have unprotected sex	-.106	.088	.078
15.	Do you currently experience symptoms of incontinence	.013	-.013	.111*
16.	I sometimes feel completely overwhelmed by all the things I need to do	-.072	-.063	.084
17.	I often feel lonely	.076	.009	.001
18.	I hate my life	.013	-.095	.009
19.	I sometimes consider committing suicide	.166	-.070	.045
20.	Are you under- or overweight	-.054	-.304*	.021
21.	I feel confident about my body	.138	-.162	.001
22.	I would like to change something about my body with plastic surgery	.058	.061	.008
23.	I sometimes avoid mirrors to prevent from having to see myself	.084	-.091	.008

[1]r values are calculated by taking the square root of the partial eta squared estimates
* $p < .05$ ** $p < .01$

4.4 Discussion

Our goal for this second study was to measure actual disclosure of information and compare the predictive power of the Rasch-model's estimate for the willingness to disclosure with that of the two currently most popular privacy instruments; the general concern for information privacy (Smith et al. 1996), and the privacy segmentation index of Kumaraguru and Cranor (2005). To measure actual disclosure of information a cover story was created about the evaluation of a new online health application. This application would provide feedback based on a set of questions about the current health situation of its users. We measured whether or not these questions were answered by our participants and whether or not they disclosed sensitive information.

Our results do not provide support for our third hypothesis, the Rasch-model's estimate for the willingness to disclose information is not a better predictor for the amount of unanswered questions. None of the privacy measures can be considered as valid predictors for the amount of unanswered questions, as indicated by the small correlations and lack of significance. Whether the participants considered the questions as odd, or difficult, and if they felt comfortable answering them were far better predictors. Our results neither provide support for our fourth hypothesis, the Rasch-model's estimate for the willingness to disclose information does not better predict the amount of sensitive information that is disclosed. The GCIP instrument performed marginally better, with one significant correlation between the instrument and a specific question, but the average correlation and its range indicate that this instrument is also not a valid predictor for the amount of disclosed information. At first glance the privacy segmentation index performed best with three significant effects, however one of these effects was in the opposite direction of what one would expect. Privacy fundamentalists disclosed most information for the question about incontinency, whereas you would expect them to disclose least information. This means we cannot consider the privacy segmentation index as a better predictor for the amount of disclosed information.

In sum, our results provide no support for our two hypothesis about the better performance of the Rasch-model estimate compared to that of the other two instruments. Nor can we say that the instrument performed worse. In general all measures can be considered as equally bad predictors for both unanswered questions as well as the amount of information disclosure, all measures showed small non-significant correlations indicating an overall poor performance.

In the next section, we will discuss both our studies in depth, and look at both the limitations and implications of our studies.

Chapter 5. General Discussion

5.1 Sensitivity of information

In this thesis we differentiated ourselves from the currently dominant view in privacy research which assumes a causal relationship between attitude and behavior. Instead we approached privacy concern from its essence, the sensitivity of personal information, and set out to answer the following research question:

Is sensitivity an invariant attribute of personal information?

In study one we adopted Campbell's Paradigm (Kaiser et al, 2010) to study the sensitivity of personal information. Our results of this study indicate that it is possible to create a list of behaviors related to information disclosure that can be transitively ordered in terms of difficulty, which is invariant in respect to individual differences. Similarly to Phelps, Nowak, and Ferrell (2000) financial information, in our case your bank balance, was considered as the most sensitive type of information. Interestingly your internet history of the past month was also found to be one of the most sensitive types of information, contradicting Carrascal, Riederer, Erramilli, Cherubini, and Oliveira's (2011) findings that offline information is more valuable to us than online information. Their sample of participants was on average a lot older, which may indicate an effect of age for this particular type of information.

The results of study one further show that we cannot ignore the influence of context, in our case the stakeholder to whom the information is disclosed. Overall Google was found to be the most difficult stakeholder to disclose information to. While this was not unexpected, commercial companies are associated with higher costs when disclosing information (Cvrcek, Kumpost, Matyas, & Danezis, 2006). It is important to realise that during the period study one took place Google received serious criticism about their privacy policies, in the form of the information about data surveillance (Greenwald & MacAskill, 2013) and their new product Google Glass (Kiss, 2013). This may have made our participants overly cautious about disclosing information to Google. While Facebook received similar critique our results indicate that Facebook has lower associated costs. This may be explained by the difference in type of questions we asked between these two stakeholders. It is possible that the self-report behavior questions for Facebook where associated with disclosing information to your Facebook friends and not to the commercial company Facebook. Alternatively there is the effect of perceived control that may explain the difference between these two stakeholders. Xu (2007) showed that perceived control lowers privacy concern, in our questionnaire Google questions were phrased passively whereas the Facebook questions were phrased as active actions. The perceived control associated with the active action of posting something on Facebook compared to the lack of perceived control when Google collects information about you may have contributed to this difference between these two stakeholders.

Most importantly our results show an interaction between information type and stakeholder, which indicates that sensitivity of personal information is not invariant of context in the form of whom the information is disclosed to. For example, the large difference between disclosing information about your medicine use to your health insurance or disclosing your holiday photo's. While both information types are similar in sensitivity, participants were far more likely to disclose their

medicine use to their health insurance than their holiday photo's. It is quite clear why this is the case, you will want to disclose your medicine use to your health insurance because you will receive immediate benefits in the form of refunds. It is however unclear what you would gain from disclosing your holiday photo's to your health insurance, let alone why they would need them. A similar argument can be made for why our participants find disclosing their internet history to Google less sensitive than disclosing it to their employer. It is quite clear why Google would like your internet history and in return you benefit in the form of getting access to their services, in the case of your employer it may be clear why he would like to know your information but it is unclear what, if at all, you get in return for the disclosure. While we can provide explanations for these effects our results indicate that we should describe information disclosure as the act of disclosing a specific type of information to a specific stakeholder.

Our first study provides an answer to our research question. Sensitivity is not an invariant attribute of personal information alone, instead it should be described as an attribute of disclosing information in a particular context. Our study further showed that sensitivity, as a combination of information and context, is independent of individual differences and was able to provide a new measure for privacy concern, the estimate for the willingness to disclose information. We will now take a closer look at this estimate and its performance compared to the two other privacy concern measures.

5.2 Willingness to disclose information

In Study one the Rasch-model's estimate for the willingness to disclose information for each individual participant was able to better predict individual pieces of information disclosure than the GCIP instrument and the privacy segmentation index. In Study two we aimed at exploring the predictive validity of this measure on actual disclosure behavior. The results of this second study do not provide support for our expectations that the Rasch-model's estimate would perform better than the GCIP and privacy segmentation index instruments. None of the instruments showed real predictive power for the disclosure of information. In this discussion we will discuss explanations for this lack of predictive power, look at limitations of the two studies, and finally the implications of our findings.

The results of our second study indicate a lack of predictive power for all three of the privacy concern measures. The explanation that first comes to mind is that we are dealing with the phenomenon of the privacy paradox. Our participants valued privacy highly as indicated by all three privacy measures and our participants disclosed large amounts of information. We see several explanations for this discrepancy between attitudes and amounts of disclosure.

First, our results indicate that we have created a scenario in our second study in which participants were willing to answer almost every question, on average only one questions was left unanswered. As a consequence our measure of disclosed information has very little variance, a so called ceiling effect, meaning we cannot expect to find correlations between this measure and the different privacy concern instruments. An explanation for this ceiling is provided by our results of Study one. Here we found that we should always consider the disclosure of information in relation to whom the information is disclosed. In our second study the stakeholder to whom the information was disclosed was a researcher at a university. Cvrcek, Kumpost, Matyas, and Danezis (2006) showed that personal information is valued as worth less when it is used for academic purposes. The large amount of answered questions and disclosed information may be explained by this academic context. This is

further emphasized by the reassurance at the start of the questionnaire that all information will be treated confidentially and will never be shared with third parties. This means that the amount of risk involved in the disclosure of the information is limited. Meanwhile the associated benefits for the participants involved with the disclosure of information are both monetary, the received incentive, and of altruistic nature, you are helping science. Thus it may be argued that we have created a context that in a Campbellian sense can be explained as a context in which information disclosure has little difficulty, as expressed in low costs, and thus a high likelihood of engagement.

Secondly, our focus within this thesis has been on the informational dimension of privacy. In our second study we have relied on information related to the health domain. The questions we asked were focused on the physical and psychological state of our participants. An argument can be made that disclosure of such information is not only part of the information dimension of privacy. If we take a closer look at the privacy dimensions as described by Burgoon et al. (1989) and Clarke (1999) it is likely that we have also tapped into physical and psychological dimensions of privacy (Burgoon et al., 1989) and the privacy of the personal body (Clarke, 1999). We have, for example, asked our participants to disclose their thoughts about suicide, this is an invasion of the psychological dimension of privacy of Burgoon et al. (1989). This possible dimensionality of our measure for information disclosure may explain the lack of correlation with instruments that are specifically focused at the dimension of information privacy.

Finally, while our participants disclosed large amounts of information and left few questions unanswered several may have followed an alternative strategy to protect their privacy. The first question in our second study asked for the name of our participants. Several choose not to answer this question, provided an alias or only used their initials, meaning they were able to fill in the questionnaire anonymously. This privacy protecting strategy is commonly used to protect privacy in social networks (Raynes-Goldie, 2010). Within our questionnaire it was the one question that would make it possible to link the answers back to them. Choosing not to answer it or providing an alias may have resulted in much lower thresholds for the other questions, as these were now answered anonymously. Our results indicate that this strategy was followed independent of individual concern, as indicated by a lack of correlation between any of the privacy measures and this specific question. Unfortunately we cannot tell if these participants would have disclosed less information if they had provided their real name, which means we cannot be sure of the influence of this privacy protecting strategy on our results. It does however illustrate that there are alternatives strategies available that assist in protecting ones privacy. And it may explain the discrepancy between privacy concern and disclosure behavior.

5.3 Limitations

There are several limitations in our two studies. In Study one we followed a facetted approach in the design of our questionnaire. Each information type was linked to each stakeholder, with the exclusion of those self-report behavior questions that were not applicable. Such an approach can quickly lead to a large amount of questions. To keep the time that participants would need to invest in our questionnaire within reasonable bounds, we limited ourselves to twelve types of information and seven stakeholders. As a consequence categories of information, for example health related information, were only presented by one item. While this is acceptable given the exploratory nature of this study it makes it difficult to draw conclusions about these types of categories. Which in turn

makes it difficult to provide practical advice to for example policy makers, here work by for example Phelps, Nowak, and Ferrell (2000) may provide more applicable insights.

Study one showed an interaction effect between context and type of information, indicating that it is important to include context in the measurement of information disclosure. We have however paid little attention to the context we created ourselves in our studies. It would for example have been valuable to have both an offline and an online version for our questionnaire, to control for an effect of answering a questionnaire online. This would also have allowed us to test our first explanation for the large amount of answered questions in study two, it would be valuable to perform Study two a second time but this in a non-academic context.

In our second study we measured information disclosure in two ways, by the amount of unanswered questions and by the amount of disclosed sensitive information. Both these measures have limitations. Most importantly we do not know if participants answered truthfully, participants may have lied to prevent information disclosure. It would have been valuable to know when they choose not to tell the truth as another measure for privacy protective behavior. Secondly in our measurement of sensitive information disclosure the vast majority of answers were only sensitive if the participants responded that the question was applicable to them, while the default situation is that it is not applicable to them. For example disclosing whether or not you have a sexually transmitted disease, is not sensitive if you do not have one which is the default situation. This means that a large amount of our questions may not have been sensitive for our participants because their response option was the non-sensitive one.

In Study two we wished to measure actual disclosure behavior, for which we made use of a cover story. We have however no measure of whether or not participants believed our cover story and were indeed unaware of the real purpose of the study. Particularly because our sample of participants was drawn from our first study in which we had explained the privacy nature of our study in the debriefing, participants may not have believed the presented cover story in the second study. If this is the case our objective measure of actual disclosure behavior will likely be biased and most likely different from real disclosure behavior.

A final limitation of our study is the relatively small sample size. A sample of 300 is still a relatively small for Rasch-model calculations. Also for the privacy segmentation index a very small set of participants was categorized as privacy fundamentalists, 8 in our second study. While the actual percentage of privacy fundamentalists was similar to other research, an univariate ANOVA with only 8 participants in one of the conditions has limited power. Especially if we consider the lack of variance in our measure for information disclosure. Real comparison between privacy concern measures requires a study with a bigger sample and with a measure that is not limited by a ceiling effect in variance of data.

5.4 Concluding remarks

In this thesis we have differentiated ourselves from the currently popular view on privacy as a causal attitude behavior relationship. Instead we adopted Campbell's paradigm with its axiomatic relationship between an attitude and a behavior. This meant returning to the essence of privacy, the sensitivity of personal information. For this we had to answer our research question: "*Is sensitivity an invariant attribute of personal information?*" In Study one it became clear that while we can describe sensitivity, as a combination of information and stakeholder, as invariant in respect to individual difference. We can however not split this difficulty into information and stakeholder as some types of information are valued completely different depending solely on whom the information is shared with. Where study one provided promising results for the estimate for the willingness to disclose information, we were unable to validate this better predictive validity in a second study. Different explanations have been provided for this lack of predictive power for our measure of the willingness to disclose information, of both methodological and theoretical nature.

With our first study we have created a first step towards a new perspective on privacy. First support was found for the application of Campbell's paradigm in privacy related research. We can order a set of behaviors about the disclosure of specific information to specific stakeholder transitively, and by using Rasch modelling create an estimate for an individual's willingness to disclose information. Our second study once more emphasized the complexity of measuring privacy and we ourselves experienced firsthand the difficulties researchers face when trying to objectively measure disclosure of information. Though we were unable to find a better performance for our measure of an individual's willingness to disclose information in our second study we do belief that Campbell's paradigm provides a promising new perspective on privacy. It is however clear that more research is needed to understand how we value the privacy of our information. Especially in today's society in which both governments and commercial companies collect more and more information about us it is imperative to know how we feel about and treat sensitive information. Only then will we as researchers be able to compare different situations with each other, and only then will we be able to provide valuable advice to policy makers and the general public to help them in their decision process about the disclosure and collection of information.

References

Acquisti, A. (2004). Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM conference on Electronic commerce* (pp. 21-29). ACM.

Acquisti, A., John, L., & Loewenstein, G. (2009). What is privacy worth. In *Workshop on Information Systems and Economics (WISE)*.

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179-211.

Bansal, G., Zahedi, F., & Gefen, D. (2010). The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems*, *49*(2), 138-150.

Bellotti, V. (1997). Design for privacy in multimedia computing and communications environments. *Technology and privacy: The new landscape*, 63-98.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences (*2nd ed.). Mahwah, NJ: Erlbaum.

Boyd, D. (2007). Social network sites: Public, private, or what. *Knowledge Tree*,*13*(1), 1-7.

Brosius, H. B., & Engel, D. (1996). The causes of third-person effects: Unrealistic optimism, impersonal impact, or generalized negative attitudes towards media influence?. *International Journal of Public Opinion Research*,*8*(2), 142-162.

Buchanan, T., Paine, C., Joinson, A. N., & Reips, U. D. (2007). Development of measures of online privacy concern and protection for use on the Internet.*Journal of the American Society for Information Science and Technology*, *58*(2), 157-165.

Burgoon, J. K., Parrott, R., Le Poire, B. A., Kelley, D. L., Walther, J. B., & Perry, D. (1989). Maintaining and restoring privacy through communication in different types of relationships. *Journal of Social and Personal Relationships*,*6*(2), 131-158.

Byrka, K. (2009). Attitude-behavior consistency: Campbell's paradigm in environmental and health domains.

Carrascal, J. P., Riederer, C., Erramilli, V., Cherubini, M., & de Oliveira, R. (2011). Your browsing behavior for a big mac: Economics of personal information online.

Clarke, R. (1999). Internet privacy concerns confirm the case for intervention. *Communications of the ACM*, *42*(2), 60-67.

Cvrcek, D., Kumpost, M., Matyas, V., & Danezis, G. (2006, October). A study on the value of location privacy. In *Proceedings of the 5th ACM workshop on Privacy in electronic society* (pp. 109-118). ACM.

Debatin, B., Lovejoy, J. P., Horn, A. K., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication*, *15*(1), 83-108.

DeCew, J. W. (1997). *In pursuit of privacy: Law, ethics, and the rise of technology*. Cornell University Press.

Fishbein, M., & Ajzen, I. (2005). The influence of attitudes on behavior. *The handbook of attitudes*, 173-221.

Greenwald, G., & MacAskill, E. (2013, June 7). NSA Prism program taps in to user data of Apple, Google and others. *The Guardian*. Retrieved September 11, 2013, from http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data.

Grossklags, J., & Acquisti, A. (2007, June). When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. In *Proceedings of the Sixth Workshop on the Economics of Information Security (WEIS 2007)* (pp. 7-8).

Hardman, D. (2009). *Judgment and decision making: Psychological perspectives* (Vol. 11). BPS Blackwell.

Harris Interactive. (2002, February). Privacy on and off the internet: What consumers want (Study No. 15229). New York, NY: Harris Interactive. Retrieved August 6, 2013, from http://www.ijsselsteijn.nl/slides/Harris.pdf

Huberman, B. A., Adar, E., & Fine, L. R. (2005). Valuating privacy. *Security & Privacy, IEEE*, *3*(5), 22-25.

Hui, K. L., Teo, H. H., & Lee, S. Y. T. (2007). The value of privacy assurance: an exploratory field experiment. *Mis Quarterly*, *31*(1), 19-33.

IBM. (1999, October). IBM multi-national consumer privacy survey. Retrieved August 6, 2013, from ftp://www6.software.ibm.com/software/security/privacy_survey_oct991.pdf

Infosecurity Europe (2004). Office Workers Give Away Passwords for a Chocolate Bar. Retrieved August 6, 2013, from http://www.net-security.org/secworld.php?id=2075

Joinson, A. N., & Paine, C. B. (2007). Self-disclosure, privacy and the Internet. *Oxford handbook of Internet psychology*, 237-252.

Kaiser, F. G., Byrka, K., & Hartig, T. (2010). Reviving Campbell's paradigm for attitude research. *Personality and Social Psychology Review*, *14*(4), 351-367.

Kaiser, F. G., Wölfing, S., & Fuhrer, U. (1999). Environmental attitude and ecological behaviour. *Journal of environmental psychology*, *19*(1), 1-19.

Kiss, J. (2013, July 2). Google Glass: privacy fears continue. *The Guardian*. Retrieved September 11, 2013, from http://www.theguardian.com/technology/2013/jul/02/google-glass-privacy-fears.

Kumaraguru, P., & Cranor, L. F. (2005). Privacy indexes: A survey of westin's studies.

Linacre, J. M. (2002). Construction of measures from many-facet data. *Journal of Applied Measurement, 3,* 486-512.

Linacre, J. M. (2003). Data variance: Explained, modeled and empirical. *Rasch Measurement Transactions, 17,* 942-943.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model.*Information Systems Research*, *15*(4), 336-355.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, *41*(1), 100-126.

Öqvist, K. L. (2009). *Virtual Shadows: Your Privacy in the Information Society*. BCS, The Chartered Institute.

Pavlou, P. A. (2011). State of the information privacy literature: where are we now and where should we go. *Mis Quarterly*, *35*(4), 977-988.

Phelps, J., Nowak, G., & Ferrell, E. (2000). Privacy concerns and consumer willingness to provide personal information. *Journal of Public Policy & Marketing*, 27-41.

Raynes-Goldie, K. (2010). Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. *First Monday*, *15*(1).

Rose, E. (2005). Data Users versus Data Subjects: Are Consumers Willing to Pay for Property Rights to Personal Information?. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on* (pp. 180c-180c). IEEE.

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS quarterly*, *35*(4), 989-1016.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information privacy: measuring individuals' concerns about organizational practices. *MIS quarterly*, 167-196.

Spiekermann, S., & Cranor, L. F. (2009). Engineering privacy. *Software Engineering, IEEE Transactions on*, *35*(1), 67-82.

Spiekermann, S., Grossklags, J., and Berendt, B. (2001) E-Privacy in Second Generation E-Commerce: Privacy Preferences versus Actual Behavior, PROC. ACM CONF. ON ELECTRONIC COMMERCE.

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, *22*(2), 254-268.

Westin, A. F (1967). *Privacy and freedom*. New York: Atheneum.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactionts, 8,* 370.

Xu, H. (2007, December). The Effects of Self-Construal and Perceived Control on Privacy Concerns. In *ICIS* (p. 125).

Appendix A: Questionnaire for Study one

On the following pages a full list of all questions asked in study one is provided, with their possible response options. The questionnaire consisted of three sections; (1) evaluative statements about the collection of information, (2) self-reported behavior questions for disclosure of information and (3) currently popular privacy concern instruments. Question order was randomized between subjects within each section. For this appendix all questions were translated from Dutch.

Evaluative Statements

Employer

	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Skip question
I accept that my employer knows my private telephone number	O	O	O	O	O	O
I accept that my employer knows my IQ	O	O	O	O	O	O
I accept that my employer knows my bank balance	O	O	O	O	O	O
I accept that my employer knows my internet history of the past month	O	O	O	O	O	O
I accept that my employer knows which party I supported at the last elections	O	O	O	O	O	O
I accept that my employer can see my holiday photo's	O	O	O	O	O	O
I accept that my employer knows which magazines I am subscribed to	O	O	O	O	O	O
I accept that my employer knows my fingerprint	O	O	O	O	O	O
I accept that my employer knows which charities I support	O	O	O	O	O	O
I accept that my employer knows which medicines I use	O	O	O	O	O	O
I accept that my employer knows my date of birth	O	O	O	O	O	O
I accept that my employer knows my highest level of education	O	O	O	O	O	O

Health Insurance

	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Skip question
I accept that my health insurance knows my private telephone number	O	O	O	O	O	O
I accept that my health insurance knows my IQ	O	O	O	O	O	O
I accept that my health insurance knows my bank balance	O	O	O	O	O	O
I accept that my health insurance knows my internet history of the past month	O	O	O	O	O	O
I accept that my health insurance knows which party I supported at the last elections	O	O	O	O	O	O
I accept that my health insurance can see my holiday photo's	O	O	O	O	O	O
I accept that my health insurance knows which magazines I am subscribed to	O	O	O	O	O	O
I accept that my health insurance knows my fingerprint	O	O	O	O	O	O
I accept that my health insurance knows which charities I support	O	O	O	O	O	O
I accept that my health insurance knows which medicines I use	O	O	O	O	O	O
I accept that my health insurance knows my date of birth	O	O	O	O	O	O
I accept that my health insurance knows my highest level of education	O	O	O	O	O	O

Municipality

	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Skip question
I accept that my municipality knows my private telephone number	O	O	O	O	O	O
I accept that my municipality knows my IQ	O	O	O	O	O	O
I accept that my municipality knows my bank balance	O	O	O	O	O	O
I accept that my municipality knows my internet history of the past month	O	O	O	O	O	O
I accept that my municipality knows which party I supported at the last elections	O	O	O	O	O	O
I accept that my municipality can see my holiday photo's	O	O	O	O	O	O
I accept that my municipality knows which magazines I am subscribed to	O	O	O	O	O	O
I accept that my municipality knows my fingerprint	O	O	O	O	O	O
I accept that my municipality knows which charities I support	O	O	O	O	O	O
I accept that my municipality knows which medicines I use	O	O	O	O	O	O
I accept that my municipality knows my date of birth	O	O	O	O	O	O
I accept that my municipality knows my highest level of education	O	O	O	O	O	O

Google

	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Skip question
I accept that Google knows my private telephone number	O	O	O	O	O	O
I accept that Google knows my IQ	O	O	O	O	O	O
I accept that Google knows my bank balance	O	O	O	O	O	O
I accept that Google knows my internet history of the past month	O	O	O	O	O	O
I accept that Google knows which party I supported at the last elections	O	O	O	O	O	O
I accept that Google can see my holiday photo's	O	O	O	O	O	O
I accept that Google knows which magazines I am subscribed to	O	O	O	O	O	O
I accept that Google knows my fingerprint	O	O	O	O	O	O
I accept that Google knows which charities I support	O	O	O	O	O	O
I accept that Google knows which medicines I use	O	O	O	O	O	O
I accept that Google knows my date of birth	O	O	O	O	O	O
I accept that Google knows my highest level of education	O	O	O	O	O	O

Self-reported behavior

Facebook

	Yes	No	Not applicable
I have shared my IQ via Facebook	O	O	O
I have shared which political party I support via Facebook	O	O	O
I have shared which magazines I am subscribed to via Facebook	O	O	O
I have shared my private telephone number via Facebook	O	O	O
I have shared my holiday photo's via Facebook	O	O	O
I have shared my highest level of education via Facebook	O	O	O
I have shared my date of birth via Facebook	O	O	O

Acquaintances

	Yes	No	Not applicable
I have shared my bank balance with my acquaintances	O	O	O
I have shared which medicines I use with my acquaintances	O	O	O
I have shared my IQ with my acquaintances	O	O	O
I have shared which political party I supported at the last elections with my acquaintances	O	O	O
I have shared which charities I support with my acquaintances	O	O	O
I have shared which magazines I am subscribed to with my acquaintances	O	O	O
I have shared my holiday photo's with my acquaintances	O	O	O
I have shared my highest level of education with my acquaintances	O	O	O
I have shared my private telephone number with my acquaintances	O	O	O
I have shared my date of birth with my acquaintances	O	O	O

Linked-In -- Evaluative statements

	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Skip question
I accept that LinkedIn knows my date of birth	O	O	O	O	O	O
I accept that LinkedIn knows which charities I support	O	O	O	O	O	O
I accept that LinkedIn knows my IQ	O	O	O	O	O	O
I accept that LinkedIn knows my private telephone number	O	O	O	O	O	O

Linked-In -- Self-reported behavior

	Yes	No	Not applicable
I have shared my date of birth via LinkedIn	O	O	O
I have shared my private telephone number via LinkedIn	O	O	O
I have shared my IQ via LinkedIn	O	O	O

General concern for information privacy (Smith et al., 1996)

	Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly Agree
It usually bothers me when companies ask me for personal information.	O	O	O	O	O	O	O
All the personal information in computer databases should be double-checked for accuracy - no matter how much this costs.	O	O	O	O	O	O	O
Companies should not use personal information for any purpose unless it has been authorized by the individuals who provided the information.	O	O	O	O	O	O	O
Companies should devote more time and effort to preventing unauthorized access to personal information.	O	O	O	O	O	O	O
When companies ask me for personal information, I sometimes think twice before providing it.	O	O	O	O	O	O	O
Companies should take more steps to make sure that the personal information in their files is accurate.	O	O	O	O	O	O	O
When people give personal information to a company for some reason, the company should never use the information for any other reason.	O	O	O	O	O	O	O
Companies. should have better procedures to correct errors in personal information.	O	O	O	O	O	O	O

	Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly Agree
Computer databases that contain personal information should be protected from unauthorized access - no matter how much it costs.	O	O	O	O	O	O	O
It bother me to give personal information to so many companies.	O	O	O	O	O	O	O
Companies should never sell the personal information in their computer databases to other companies.	O	O	O	O	O	O	O
Companies should devote more time and effort to verifying the accuracy of the personal information in their databases.	O	O	O	O	O	O	O
Companies should never share personal information with other companies unless it has been authorized by the individuals who provided the information.	O	O	O	O	O	O	O
Companies should take more steps to make sure that unauthorized people cannot access personal information in their computers.	O	O	O	O	O	O	O
I'm concerned that companies are collecting too much personal information about me.	O	O	O	O	O	O	O

Privacy Segmentation Index (Kumaraguru & Cranor, 2005)

	Strongly disagree	Somewhat disagree	Somewhat agree	Strongly agree
Consumers have lost all control over how personal information is collected and used by companies	O	O	O	O
Most businesses handle the personal information they collect about consumers in a proper and confidential way	O	O	O	O
Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today	O	O	O	O

Appendix B: Questionnaire for Study two

On the following pages a full list of all questions in study two is provided. All text is translated from Dutch.

General Questions

1. What is your name:

_____

2. What is your age:

_____

3. What is your gender:

- **O** Male
- **O** Female
- **O** Skip this question

4. To which ethnic group do you belong?

_____

5. Are you religious

- **O** Catholic
- **O** Dutch reformed church
- **O** Calvinistic
- **O** Islamic
- **O** Other
- **O** Not religious
- **O** Skip this question

6. Do you currently user any medicines?

- **O** Yes
- **O** No
- **O** Skip this question

7. Are there any genetic diseases in your family?

- **O** Yes, and I am diagnosed
- **O** Yes, and I am at risk
- **O** Yes, but I am not at risk
- **O** No
- **O** Skip this question

Dietary Questions

8. I make sure I eat healthy

- **O** Agree
- **O** Disagree
- **O** Skip this question

9. I eat breakfast every day

- **O** Yes
- **O** No
- **O** Skip this question

10. I am an emotional eater

- **O** Agree
- **O** Disagree
- **O** Skip this question

11. I sometimes feel guilty after eating something unhealthy

- **O** Agree
- **O** Disagree
- **O** Skip this question

12. On how many days during a week do you eat meat or fish produce?

_____

Alcohol and other Drug related questions

13. How many glasses of alcohol do you consume in an average week?

_____

14. I have once drunk so much that I could not remember everything the day after

- **O** Agree
- **O** Disagree
- **O** Skip this question

15. I sometimes drink alcohol because it fits in the social situation and not because I want to

- **O** Agree
- **O** Disagree
- **O** Skip this question

16. I have done something when I was drunk that embarrasses me

- **O** Agree
- **O** Disagree
- **O** Skip this question

17. Do you use soft-drugs?

- **O** Yes
- **O** No, but I have tried them
- **O** No
- **O** Skip this question

18. Do you use hard-drugs?

- **O** Yes
- **O** No, but I have tried them
- **O** No
- **O** Skip this question

Sexual activity

19. Are you currently in a relationship?

- **O** Yes
- **O** No
- **O** Skip this question

20. Have you ever cheated on someone?

- **O** Yes, multiple times
- **O** Yes
- **O** No
- **O** Skip this question

21. What is your sexual orientation?

- **O** Heterosexual
- **O** Homosexual
- **O** Bisexual
- **O** Asexual
- **O** Skip this question

22. Are you currently sexually active?

- **O** Yes
- **O** No
- **O** Skip this question

23. With how many people have you had sex in the past year?

- **O** 0
- **O** 1-2
- **O** 3-4
- **O** 5-6
- **O** 7 or more
- **O** Skip this question

24. Have you ever had a sexually transmittable disease (STD)?

- **O** Yes, and still do
- **O** Yes
- **O** No
- **O** Skip this question

25. I have had sex with more than one person within a day (24 hours)

- **O** Yes
- **O** No
- **O** Skip this question

26. I sometimes have unprotected sex

- **O** Yes
- **O** No
- **O** Skip this question

27. In case of unwanted pregnancy an abortion is a serious consideration for me

- **O** Agree
- **O** Disagree
- **O** Skip this question

28. Do you currently experience symptoms of incontinence

- **O** Yes
- **O** No
- **O** Skip this question

Questions related to self-image

29. I sometimes feel completely overwhelmed by all the things I need to do

- **O** Agree
- **O** Disagree
- **O** Skip this question

30. I am often tired and feel a lack of energy

- **O** Agree
- **O** Disagree
- **O** Skip this question

31. I am often tired and feel a lack of energy

- **O** Agree
- **O** Disagree
- **O** Skip this question

32. I often feel lonely

- **O** Agree
- **O** Disagree
- **O** Skip this question

33. I hate my life

- **O** Agree
- **O** Disagree
- **O** Skip this question

34. I sometimes consider committing suicide

- **O** Agree
- **O** Disagree
- **O** Skip this question

35. I feel confident about my body

- **O** Agree
- **O** Disagree
- **O** Skip this question

36. I would like to change something about my body with plastic surgery

- **O** Agree
- **O** Disagree
- **O** Skip this question

37. I sometimes avoid mirrors to prevent from having to see myself

- **O** Agree
- **O** Disagree
- **O** Skip this question

Evaluation of Total Fit

	Disagree	Slightly disagree	Neutral	Slightly agree	Agree
The presented tips were useful for me	O	O	O	O	O
The presented tips were complete	O	O	O	O	O
The presented tips will help me achieve a healthier lifestyle	O	O	O	O	O
I enjoyed using the health application	O	O	O	O	O
I have used a similar application before	O	O	O	O	O
I would like to make use of the Total Fit application	O	O	O	O	O
I did not find it difficult to answer the questions	O	O	O	O	O
I considered the questions odd	O	O	O	O	O
I felt comfortable answering the questions	O	O	O	O	O