BEYOND AUTONOMY AND CONTROL: AN IDENTITY-BASED VIEW OF INFORMATION PRIVACY ON SOCIAL MEDIA

Research in Progress

Wu, Philip Fei, Royal Holloway, University of London, Egham, UK, philip.wu@rhul.ac.uk

Abstract

This paper proposes an identity-based conceptualization of information privacy and suggests that privacy should be understood against the backdrop of self-identity formation. The researcher argues that the so-called "privacy paradox" - the baffling contradiction between grave privacy concerns in society and the prevalence of information sharing on social media - is not a paradox per se; rather, privacy concerns reflect the ideology of autonomous self, whereas social construction of identity explains voluntary information disclosure. The researcher first unpacks the mainstream conception of autonomy-centric privacy in the IS literature, and then present a research model that illustrates a theorization of the relationship between privacy and self-identity. The paper also reports a pilot study that validates the constructs for future research.

Keywords: Privacy, Autonomy, Identity, Social media.
1 Introduction

As people leave more digital footprints in this highly connected world, privacy has become the issue of our times (Acquisti et al., 2015). Most IS scholars define privacy based on Altman’s (1975) notion of privacy as “the selective control of access to the self” (p.18) and Westin’s dichotomy of “self versus society” (Conger et al., 2013; Dinev et al., 2013). Thus, privacy is studied from the perspective of “privacy calculus” where the core principle is the control of informational transactions between an individual and others, and “the ultimate aim of which is to enhance autonomy and/or to minimize vulnerability” (Margulis 1977, cited in Smith et al. 2011, p. 995). It is perhaps the emphasis on vulnerability that has led IS researchers to focus on perceptual privacy concerns, whereas the effort of theorizing privacy as a social and psychological concept has been limited (Dinev et al., 2013).

This paper argues that privacy should also be understood against the backdrop of identity formation, which is an on-going process characterized by social interactions and information disclosure. If the self is an intersubjective entity being constructed in the presence of others (Jenkins, 2008; Mead, 1967), the privacy concerns of protecting “the self” only represent one side of the story. The “privacy paradox” – the increasing awareness of privacy issues coincides with the omnipresent information sharing on the Internet – reveals the limitation of this protection narrative (Acquisti et al., 2015; Brandimarte et al., 2013; Taddicken, 2014). The other side of the story, the researcher contends, lies in the fact that individuals form their self-identities through disclosing “the self” in social interactions.

A few recent papers authored by industry researchers have already advocated privacy management approaches that go “beyond access control” (Krishnamurthy, 2013; Mondal et al., 2014). Social media companies have also begun to recognize privacy needs in different social scenarios and tweaked their platform designs in recent years. For example, Google+ Circles and Facebook Groups allow users to categorize their connections based on social categories (acquaintances, close friends, co-workers, etc.). An individual’s privacy valuation in each of these circles or groups would be different. In academic research, Squicciarini et al.’s (2011) idea of collaborative privacy management on social network sites is a step closer to identity-based privacy management, although the study still clings on the access control paradigm in describing information “ownership” and “stakeholders”. Schwaig et al. (2013) considered identity-related concepts such as consumer alienation and self-esteem, but they are conceptualized as “individual differences” rather than key theoretical constructs.

In contrast to the autonomy-centric view of privacy, this paper proposes an identity-based conceptualization of privacy which highlights the interdependence between privacy and information disclosure. The researcher also presents in this paper a research model and reports a pilot study of validating the constructs. This paper makes three contributions. Firstly, it helps to resolve the so-called “privacy paradox” by providing a theory-informed explanation to the phenomenon: both information sharing and privacy management are for the same purpose of constructing a self-identity. Secondly, drawing on identity research in social psychology, the paper broadens the understanding of privacy as a sociopsychological construct. The dialectic relationship between protecting and constructing the self in social interactions suggests new possibilities in theorizing privacy in today’s media-rich information environment. Thirdly, to the best of the researcher’s knowledge, this is one of the first studies exploring the connections between self-identity, privacy, and online information disclosure.

2 Privacy and Autonomy

The autonomy-centric perspective has helped to explain a great deal of people’s privacy perceptions and behaviors on social media (for reviews of IS literature on privacy, see Bélanger and Crossler, 2011; Smith et al., 2011). As information disclosure is often necessary for a consumer to receive commercial services, many have argued that privacy is essentially a trade-off between autonomy and
self-disclosure (Krasnova et al., 2010; Pavlou, 2011). What is puzzling, however, is that past research has found little evidence of correlation between privacy concerns and the overall online self-disclosure behavior (Taddicken, 2014). In other words, people seem to act on a more complex set of mechanisms than the simple trade-off principle. Two underlying assumptions in this trade-off argument deserve scrutiny: 1) Human beings are rational animals who make autonomous decisions based on risk-benefit assessment; 2) Privacy is a static valuation criterion against which individuals assess the risks of information disclosure.

The first assumption of Homo Economicus has been subject to much criticism from within and outside the management discipline and the researcher shall not repeat the argument in this short paper. For a review in the management literature, see Folger and Salvador (2008). The second assumption implies a fixed privacy perception and a mechanical decision-making process, both of which are questionable beliefs. Using the analogy of architecture versus archaeology, Bettman et al. (1998) argue that consumer preferences are constructed (as architecture) rather than uncovered (as archaeology). The constructive process is contingent upon information environment, limited by bounded rationality (Simon, 1955), and shaped by interactions between human information processing and feedback. In reality, privacy preferences are even more complex than purchase decisions, which are usually based on factual product attributes (e.g., elimination-by-aspects, weighted adding, etc.) are not very useful.

The “trade-off” argument in the privacy literature provides a convenient justification for information disclosure but it does not adequately explain the psychological mechanisms of making the trade-off choices (Bettman et al., 1998). In fact, past research has shown that people can be uncertain about their own privacy preferences even when they are aware of the consequences of their privacy decisions. For instance, in a series of experiments, Brandimarte et al. (2013) demonstrated that giving people more control over the publication of their personal information decreases their privacy concerns and increases their willingness to share, even when the risk probabilities remain the same or even increase. Stutzman et al. (2013) analyzed a longitudinal panel of 5,076 Facebook users’ profile data over six years and found that information disclosure in connected private profiles intensified despite the increasingly privacy-seeking behavior observed in the public network. Thus, privacy management on Facebook actually led to more information disclosures to “silent listeners” including Facebook itself and advertisers, often without the users’ consent or awareness.

The view of a static, context-independent privacy valuation is also reflected in Westin’s (1967) famous taxonomy of individual preferences of privacy: privacy fundamentalists, pragmatists, and unconcerned. In today’s intellectual context, these categories are somewhat problematic. Firstly, the term “fundamentalists” carries such a negative connotation that it seems inappropriate to label those who hold dear their privacy rights. Secondly, when an individual appears “unconcerned” by privacy, it is often because s/he is under-informed or ignorant of privacy violations by businesses and governments (Cohen, 2012; Lessig, 2006). For instance, a 2012 Pew report found that social media users with lower levels of education are significantly less likely to report privacy concerns than those with college education (Madden, 2012). Thirdly, many studies reported a discrepancy between people’s privacy attitudes and their actual behaviors. Tufekci (2008) found little or no association between college students’ online privacy concerns and their information disclosure behaviors. In a large-scale laboratory experiment, Berendt (2005) found that even the self-claimed “privacy fundamentalists” forgot their privacy concerns in a rich interactive shopping environment. Similarly, in a ubiquitous computing experimental setting, Connelly et al. (2007) measured participants’ privacy concerns using both a survey and in-situ questionnaires and found less than a third of the answers from the two questionnaires were identical.

All these studies seem to point to a dull conclusion: we are all “privacy pragmatists”. People are willing to trade their autonomy in exchange for goods and services, despite privacy concerns and risks. In the context of using social media, much of the past IS research made the observation that individuals value the convenience of maintaining social relationships in online social networks. But why main-
taining social relationships online is so important, to the extent that people ignore privacy risks? Why is there a discrepancy between self-reported privacy concerns and actual privacy practices in online spaces? The “trade-off” or “privacy as a commodity” perspective may not be able to fully explain the seemingly contradictory information behaviors. Some researchers attempt to draw upon Social Exchange theory to examine the willingness of self-disclosure (e.g., Krasnova et al., 2010). However, social exchange theory is about micro social orders that are sustained by the principle of reciprocity; it has little to do with the exchange of information as goods between individuals and corporates (for reviews of social exchange theory, see Cropanzano and Mitchell, 2005; Zafirovski, 2005). There are deep psychological reasons to people’s “pragmatist” privacy behaviors.

3 An Identity-Based View of Online Privacy

Identity theorists\(^1\) regard the self as a reflective existence that can categorize itself in relation to other social categories (Stets and Burke, 2000). Through this process of self-categorization (a vocabulary of social identity theory) or identification (a vocabulary of identity theory), self-identity is formed. Because social categories for each individual vary, identity theorists place a great emphasis on the interconnected individuality in interaction contexts. Therefore, the self does not merely exist at the level of one’s unique individuality (as is usually assumed in personality psychology); rather, the self is always an intersubjective entity that implies the presence of others (Jenkins, 2008; Mead, 1967). Through a cyclical and negotiating process, individuals come to find self-understanding and be able to express the self-identity (Floridi, 2011).

Thus, although individuals attempt to distinguish themselves from their interaction counterparts, the dissimilarities must be performed and then negotiated in social interactions. As Cohen (2012) puts it, citing performance theorists, “identity in a social world exists only insofar as it is performed to and for others” (p. 130). Evidences of identity performance in online social networks have been documented in recent literature. Pluempavarn and Panteli (2008) described how individual bloggers construct their social identity through shaping the collective identity of the blogging community. boyd and Heer (2006) analyzed millions of Friendster profiles to explore how users perform and communicate their identities through crafting their profiles. Those online profiles are not just embodied identities but invitations for “communicative dances” with both known and imagined audiences. Similar self-presentation and audience management strategies are seen in Twitter networks, where users maintain their “authentic me” through self-censoring their tweets (Marwick and boyd, 2011).

If one accepts that self-identity is fundamentally social, it is then immediately obvious that information disclosure is integral to privacy. Individuals are never truly autonomous and privacy is not only about information protection. In a world of mass production and all-pervading commodification, the physical objects we possess are mostly reproducible and identical to what others have. Sociologists believe that anxiety may arise from being unable to discern “self” from “others” in society (Giddens, 1991). This has led to, in the words of Floridi (2010), rampant “informational re-appropriation” (p.15) in online spaces: we share information about ourselves to become less informationally indiscernible. In other words, people try to retain individualism by giving away individual details.

While some researchers view the desire for disclosure and the need for privacy as two competing motivations (Acquisti et al., 2015), few have elaborated the relationship between them from the angle of self-identity. This paper proposes three premises that capture the dynamics between self-identity, privacy management behavior, and information disclosure.

\(^1\) Identity theory and social identity theory have roots in different intellectual traditions, yet these two theories have substantial overlap. See Hogg et al. (1995) and Stets and Burke (2000).
Premise 1: The need for self-identity is positively related to information disclosure. The need for self-identity refers to the psychological need to define and evaluate oneself in social life (Schlenker 1982; Pierce et al. 2001). Social psychology literature tends to focus on two aspects of the need: the need of self-awareness and the need to communicate self-identity to others (Jenkins 2008; Pierce et al. 2003). Self-awareness is “what the individual is conscious ‘of’ in the term ‘self-consciousness’” (Giddens, 1991, p.53) and answers the question – “Who am I?” On the other hand, the coherent self-consciousness must be validated and constantly adjusted through communicating with others (Jenkins 2008). People feel understood and satisfied when their self-presented identities are confirmed in social interactions (Goffman, 1967; Swann and Read, 1981). Evidences from IS-related disciplines (HCI, Communications) support the potential linkage between the need for self-identity and personal information disclosure in online social interactions. In a widely cited paper, Ellison et al. (2007) studied college students’ Facebook usage and described a strong connection between students’ self-esteem and intensity of using Facebook. Child et al. (2009) found that online bloggers’ self-consciousness was positively related to open disclosure: individuals with higher levels of internal self-consciousness also were more likely to enact public blogging and, subsequently, privacy management practices. Interpersonal communication research points out that self-disclosure and some form of vulnerability are necessary in developing deeper intimacy with another person (Altman and Taylor, 1973). Ma and Agarwal (2007) analyzed how technology artifacts in online communities afford identity communication and verification in the process of knowledge sharing. Bumgarner (2007) and Joinson (2008) both described Facebook as a place for establishing shared identities through exhibitionism, gossip, and virtual people watching.

Based on Premise 1 and prior studies on social media, the researcher derives two operational hypotheses for the present study:

Hypothesis 1a: The more an individual is conscious of his or her self-identity, the more likely s/he will disclose personal information on social media.

Hypothesis 1b: The more an individual feels the need for expressing his or her self-identity, the more likely s/he will disclose personal information on social media.

Premise 2: The need for self-identity is positively related to privacy management behavior. Dourish and Anderson (2006) argue that privacy is a collective (rather than individual) information practice being enacted to demarcate social boundaries – the boundaries between “us” and “them”. The authors cite studies of teenagers’ secret-keeping behaviors, amateur mushroom enthusiasts’ group interactions, and long-haul truckers’ information behaviors to illustrate that privacy behavior is a marker of social affiliation and group identity. The present study takes this argument one step further by postulating that the very purpose of privacy management is to satisfy the need for self-identity. In the context of online interaction, privacy management includes activities such as tweaking privacy settings on social networking sites, pruning online personal profiles, removing digital footprints, and so forth (Madden, 2012). Since people engage in the identification process through mechanisms such as self-representation and identity verification, the researcher postulate that privacy management is to facilitate self-presentation (boyd and Heer, 2006), manipulate perceived identity verification (Ma and Agarwal, 2007), and define the parameters of social comparison (Lee, 2014).

Two operational hypotheses may derived from Premise 2:

Hypotheses 2a: The more an individual is conscious of his or her self-identity, the more likely s/he will engage in privacy management on social media.

Hypothesis 2b: The more an individual feels the need for expressing his or her self-identity, the more likely s/he will engage in privacy management on social media.

Premise 3: Privacy management is positively related to information disclosure. Lewis (2011) examined the evolution of privacy behavior on Facebook over four years and found an interdependence be-
tween friendship decisions and privacy behavior: on one hand, more college students chose to have a private profile over time, but each individual’s network size also increased; on the other hand, students with larger networks are more likely to have a private profile. In other words, there seem to be a strangely positive correlation between the act of keeping things private and the act of making more friends. Another longitudinal study on Facebook by Stutzman et al. (2013) reported a very similar observation and provided a plausible explanation: access to increasingly granular privacy settings on Facebook increases users’ feeling of control and encourages high level of information disclosure.

These observations corroborate with findings from another stream of privacy research that focused on effectiveness of privacy management mechanisms on social media. Combining survey and Facebook log data analysis, Bernstein et al. (2013) discovered that the users’ estimation of their Facebook audience was only 27% of its true size. This mismatch, they argue, might have encouraged more information sharing as some users are uncomfortable broadcasting to a large audience. Liu et al. (2011) found that in their Facebook sample the privacy settings match users’ expectations only 37% of the time, and when incorrectly configured, almost always result in more information disclosure to unexpected audience. Mondal et al.’s (2014) recent study on social access control lists (SACLs) (e.g., Facebook “Friends Lists” and Google+ “Circles”) revealed the complexity of identifying subsets of friends when sharing, which questions the extent to which those SACLs capture the users’ real privacy preferences. In short, the privacy mechanisms implemented by social media companies often gave users the “illusory” control (or lack thereof) (Hoadley et al., 2010) over private information whereas the actual risks are not necessarily mitigated.

Therefore, following Stutzman et al. (2013), the researcher postulates that privacy management tools on social media are likely to create a sense of control and safety, which will encourage users to share more personal information:

**Hypothesis 3:** The more an individual engages in privacy management when using social media, the more likely s/he will disclose personal information on social media.

From an autonomy-centric perspective, the positive association between privacy management and information disclosure may sound counter-intuitive. But the empirical evidences in the cited literature illustrate the conception of the dialectic relationship between privacy and identity: privacy is for protecting the self but at the same time also for constructing self-identity through disclosing the self to others. The relationships among identity, privacy, and information disclosure may be illustrated in a research model shown in Figure 1.

![Research model](image)

**Figure 1.** Research model

## 4 Pilot Study

The hypotheses will be tested with a survey method. The researcher searched the IS, psychology, and HCI literatures to identify rigorously validated empirical instruments. Fenigstei et al.’s (1975) Self-Consciousness Scale and the revised version by Scheier and Carver (1985) have been widely used in
psychology research. The self-consciousness scale was also adopted in some online privacy studies such as Child et al. (2009) and Lee (2014). The scale contains different sets of items for private and public self-consciousness. Private self-consciousness concerns attending to one’s inner thoughts and feelings, while public self-consciousness is about “a general awareness of the self as a social object” (Fenigstein et al., 1975, p. 523). The researcher used the seven public self-consciousness items in the survey. Survey respondents were asked to indicate how much the statement in each item “is like you” using a scale of 1 (“not like me at all”) to 4 (“a lot like me”).

The psychological need for expressing one’s self-identity has been well documented in social sciences, especially in the impression management literature (e.g., Leary, 1996). However, most prior studies either did not take a survey approach, or measured needs satisfaction (e.g., Ellison et al., 2007) or the self-presentation behaviors for fulfilling the need (e.g., Ma and Agarwal, 2007), rather than the need per se. Hence, a new scale that measures the psychological need for self-expression had to be developed. In the search for relevant publications, the researcher found theories and concepts of self-presentation particularly pertinent (Goffman, 1967; Jones and Pittman, 1982; Leary, 1996). The researcher also reviewed IS and HCI publications that applied the self-presentation concept (e.g., boyd and Heer, 2006; Dijck, 2013; Forman et al., 2008; Joinson, 2001; Ma and Agarwal, 2007; Marwick and boyd, 2011; Tufekci, 2008). In addition, the researcher found many studies in the organizational psychology literature that address employee’s self-identity. For example, Pierce et al.’s (2001, 2003) psychological ownership theory posits that the need for self-identity is an integral dimension of psychological ownership as people seek opportunities to express their identity through owning tangible and intangible objects. The researcher developed an initial set of six items that represented the latent construct of need for self-expression.

For online information disclosure behavior, most of the items were adapted from Taddicken’s (2014) Self-Disclosure on the Social Web scale. Survey participants responded to the question “Which information are you revealing on social media?” using a 4-point scale for each item, with 1 = “Never”, 2 = “Rarely”, 3 = “Frequently”, 4 = “Always”. To gauge privacy management practices, the researcher created five items based on (2009) survey on Facebook users’ privacy practices and Madden’s (2012) Pew Internet report on privacy management on social media sites. The scale included statements about common privacy control behaviors across social media platforms, such as adjusting default privacy settings and removing privacy sensitive information on social media. Respondents were asked to rate the extent to which the statements are applicable to them.

Following the guidelines by Straub et al. (2004), the researcher assessed the survey instrumentation’s content validity, construct validity, and reliability. Content validity is usually established through literature review and domain expert review. In this case, most of the survey items were adopted from previously validated instruments in the literature, and an early version of the questionnaire was sent to four senior academics for review. Additional feedback was gathered through a free-text question at the end of the online pilot study (detailed below). After carefully considering both the experts’ and the respondents’ suggestions, the researcher made changes to the instrument, including rephrasing certain items, reordering the blocks of questions, and improving the online interface.

To assess the instrumentation’s construct validity and reliability, the researcher conducted a pilot study with workers on Amazon Mechanical Turk (MTurk). MTurk is an online labor marketplace where registered workers volunteer to perform small tasks (called Human Intelligence Tasks or HITs) for micro payments. Past research has shown that MTurk has the advantage of reaching a more diversified research population than college students and the quality of data collected on MTurk are as good as, if

---

2 Scheier and Carver’s (1985) original scale uses a 0-3 Likert scale where 0 = “not like me at all” and 3 = “a lot like me”. For the sake of consistency with other scales used in this study, the researcher used a 1-4 scale.
not better than, that collected in conventional environments (Buhrmester et al., 2011; Mason and Suri, 2012; Paolacci et al., 2010). In recent years, IS researchers have also used MTurk to collect empirical data with success (e.g., O’Leary et al., 2014; Tang et al., 2015; Wu, 2013). To ensure the quality of the data, the researcher set the prescreening criteria on MTurk to restrict survey access to workers who had a high HIT approval rate (greater than 95%), lived in the United States, and were active social media users. Once a worker accepted the HIT, he or she was then directed to a Web-based survey platform (Qualtrics) where the survey was hosted. Qualtrics was also configured to allow only one survey response from each IP address and the IP must be located in the US. Upon completing the questionnaire on Qualtrics, each respondent received a unique code and was instructed to enter the code on MTurk to claim the payment. 168 MTurk workers accepted the HIT and 165 responses were complete and usable.

The researcher built a measurement model using SmartPLS (Ringle et al., 2015) and conducted confirmatory factor analysis to assess the validity (convergent validity and discriminant validity) and reliability of the instrument (Gefen and Straub, 2005). Convergent validity is the degree to which the measurement items for a theoretical construct are correlated with one another, whereas discriminant validity is the degree to which the measures of each construct differ (Straub et al., 2004). Upon examining the measurement items’ factor loadings and cross-loadings on each construct, the researcher dropped one item from the Self-Expression scale and three items from the Self-Consciousness scale due to their low loadings (<.65). All other factor loadings were greater than .70 with no cross-loadings above .50. The average variances (AVEs) extracted for the constructs ranged from .58 to .79, demonstrating a good convergent validity of the measurement model; the square roots of the AVEs were larger than any correlation between the constructs, which supported the discriminant validity. Reliability is usually assessed by two criteria: Cronbach’s alpha and composite reliability. In this case, Cronbach’s alphas ranged from .83 to .93 and the composite reliability .88 to .95, both indicating excellent reliability of the instrumentation. In summary, the pilot study established the validity and reliability of the construct measures. The next step of this research is to collect a larger sample of data to assess the structural model for hypothesis testing.

5 Conclusion

This paper conceives privacy in relation to identity rather than autonomy. This is not a mere word play. Autonomy is a value-laden term that carries the baggage of liberal individualism (Cohen, 2012), which may overshadow alternative conceptualizations of privacy in academic discussions. Identity, on the other hand, is a more neutral concept that invites diverse discourses from various intellectual traditions. Moving beyond autonomy-centric privacy discussions helps to deepen the IS community’s understanding of privacy and avoid pitfalls in what Smith et al. (2011) called “normative and sometimes emotionally charged” (p.1003) privacy studies.

The researcher is not claiming that an identity-based view is superior to the autonomy-centric one. Autonomy and information control are important aspects of privacy, especially in e-commerce environments where data tracking is omnipresent. In business transactions, consumers are perhaps more concerned about protecting identifiable personal information than establishing social identities. However, privacy needs in online social interactions are slightly different from those in economic exchanges. Digital technologies have the affordances of constructing a virtual presence through which the self seeks to identify itself in a potentially feedback loop. An identity-based conception of privacy, therefore, solves the so-called “privacy paradox” - the baffling contradiction between grave privacy concerns in society and the prevalence of information sharing on social media. The researcher argues that it is not a paradox per se: privacy concerns reflect the ideology of autonomous self, whereas social construction of identity explains voluntary information disclosure. While Figure 1 presents a possible research model and it is still subject to empirical validation, the researcher hopes the identity-based theorization helps to inspire novel ideas in future privacy research.
References


