Communications of the Association for Information Systems

Volume 41 Article 4

8-2017

Familiarity with Big Data, Privacy Concerns, and Self-disclosure Accuracy in Social Networking Websites: An APCO Model

Tawfiq Alashoor
Georgia State University, talashoor1@gsu.edu

Sehee Han
Penn State Harrisburg

Rhoda C. Joseph
Penn State Harrisburg

Recommended Citation
Alashoor, Tawfiq; Han, Sehee; and Joseph, Rhoda C. (2017) "Familiarity with Big Data, Privacy Concerns, and Self-disclosure Accuracy in Social Networking Websites: An APCO Model," *Communications of the Association for Information Systems*: Vol. 41 , Article 4.
DOI: 10.17705/1CAIS.04104
Available at: https://aisel.aisnet.org/cais/vol41/iss1/4

Communications of the Association for Information Systems

Research Paper ISSN: 1529-3181

Familiarity with Big Data, Privacy Concerns, and Self-disclosure Accuracy in Social Networking Websites: An APCO Model

Tawfiq Alashoor
Department of CIS, Georgia State University
Department of Accounting and MIS, King Fahd University of Petroleum and Minerals
talashoor1@gsu.edu

Sehee Han
Department of Public Administration
Penn State Harrisburg

Rhoda C. Joseph
School of Business Administration
Penn State Harrisburg

Abstract:

Social networking websites have not only become the most prevalent communication tools in today's digital age but also one of the top big data sources. Big data advocates promote the promising benefits of big data applications to both users and practitioners. However, public polls show evidence of heightened privacy concerns among Internet and social media users. We review the privacy literature based on protection motivation theory and the theory of planned behavior to develop an APCO model that incorporates novel factors that reflect users' familiarity with big data. Our results, which we obtained from using a cross-sectional survey design and structural equation modeling (SEM) techniques, support most of our proposed hypotheses. Specifically, we found that that awareness of big data had a negative impact on and awareness of big data implications had a positive impact on privacy concerns. In turn, privacy concerns impacted self-disclosure concerns positively and self-disclosure accuracy negatively. We also considered other antecedents of privacy concerns and tested other alternative models to examine the mediating role of privacy concerns, to control for demographic variables, and to investigate different roles of the trust construct. Finally, we discuss the results of our findings and the theoretical and practical implications.

Keywords: Social Networking Websites, Privacy Concerns, APCO, Big Data, Disclosure Outcomes, SEM.

This manuscript underwent editorial review. It was received 01/13/2016 and was with the authors for 14 months for 2 revisions. Paul Benjamin Lowry served as Associate Editor.

1 Introduction

The number of social networking websites and their users continues to grow. According to recent polls, 52 percent of American adults use two or more social networking websites (Duggan, Ellison, Lampe, Lenhart, & Madden, 2015). Facebook still dominates the social media industry: it counts 71 percent of Americans as users (Duggan et al., 2015). Other social networking websites (e.g., LinkedIn, Pinterest, Instagram, and Twitter) share a relatively similar number of users that ranges from 23 to 28 percent of Americans (Duggan et al., 2015). Social networking websites—Internet communities where individuals communicate and exchange digital information through personal profiles—have become the most prevalent communication tools in today's digital age (Acquisti & Gross, 2006; Boyd, 2004). Social networking websites enable individuals to extend beyond physical boundaries and to connect with billions of users (Cao, Basoglu, Sheng, & Lowry, 2015). As a result, researchers have predicted that they will contribute to the already exponentially growing rate of digital data production and that they will have a strong impact on businesses and societies (Cao et al., 2015; Gantz et al., 2008). However, the fact that various parties purposely collect, permanently store, and process in various ways the information shared on these websites raises several concerns. In this regard, protecting the privacy of personal information is one of the most challenging concerns for both service providers and users (Boyd & Crawford, 2012; Breznitz, Murphee, & Goodman, 2011; Chan et al., 2005; Shim, Dekleva, French, & Guo, 2013).

Public polls show evidence of heightened privacy concerns among Internet and social media users (GfK, 2014; Rainie, Kiesler, Kang, & Madden, 2013; Duggan et al., 2015). One can heuristically predict this growing concern as we consider the reality of today's digital age. First, it is easier than ever to look up someone's personal information. Many websites and software programs exist (some free of charge) to provide interested people with personal data about social networking website users (Breznitz et al., 2011; Craig & Ludloff, 2011; He, Zha, & Li, 2013). Second, the number of data breach incidents continues to increase (Shey, 2013; Ponemon, 2014a, 2014b). Third, practitioners do not want to miss the lucrative opportunities social networking websites provide. Marketers tailor their ads by processing users' personal data, data brokers make profits from users' personal data, employers can easily gain access to personal profiles, and governments monitor social media all the time (Breznitz et al., 2011; Drake, Hall, Becton, & Posey, 2016; Hurwitz, Nugent, Halper, & Kaufman, 2013; Schmarzo, 2013). Social networking websites operate giant databases that contain a massive amount of collected private data. Accordingly, when malicious individuals breach or easily obtain the data that these databases store, they put individuals in jeopardy by potentially exposing them to identity theft, embarrassment, and other threats (Choi, Jiang, Xiao, & Kim, 2015; Watson, 2014). Add to this point the ramifications of limited regulations and weak privacy policies that social networking websites adopt (Acquisti & Gross, 2006; Chen & Rea, 2004; Gundecha & Liu, 2012; Fernback & Papacharissi, 2007), one can see why public polls and empirical research have found high levels of privacy concerns.

The above stated realities about today's digital age reflect an era of big data where individuals' personal data are the driving force (Watson, 2014). Researchers have predicted big data and big data analytics to lead the information technology (IT) industry in the coming few years (Manyika et al., 2011). Big data refers to "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyika et al., 2011, p. 11). Recently, Facebook, Twitter, and other well-known social networking websites have adopted big data technologies. Now, social networking websites have powerful capabilities to develop "digital dossiers at a level of detail that we have never seen before" (Shim, French, Guo, & Jablonski, 2015, pp. 39). They can use these dossiers (i.e., a collection of personal data) for various purposes including basic and advanced analytics for insight, operational analytics, and monetized analytics (for more information, see Hurwitz et al., 2013). In this study, we investigate two important issues in the intertwined contexts of social networking websites and big data.

First, Watson (2014, pp. 1263) suggests that users are likely to become more concerned as they become familiar with the fact that companies are increasingly using big data analytics. However, no empirical evidence supports this assertion. In other words, we do not know whether familiarity with the concept and practices of big data impacts users' attitudes toward privacy and disclosure outcomes. Therefore, we introduce the construct of familiarity with big data and empirically test how it impacts privacy-related constructs.

Second, while more organizations have increasingly begun to rely on social media as a big data source, insights from this data source come with an array of credibility and reliability issues (Abbasi, Sarker, & Chiang, 2016; Keith, Babb, & Lowry, 2014; Trinkle, Crossler, & Bélanger, 2015) or what Hurwitz et al.

(2013) refer to as validity and veracity issues. We maintain that one can partially attribute these issues to the accuracy of the data that users originally reveal. For instance, users of social networking websites are likely to pursue some protective behaviors (e.g., falsifying personal information) because of their concerns and doubts "on what big data companies such as Facebook and Google can do with the data they collect" (Watson, 2014, pp. 1248). A high level of inaccurate and incomplete information in a big data set extracted from a social networking website results in low reliability. Consequently, low reliability has a profound negative impact on the veracity of big data analytics and, hence, undermines the overall utility of big data analytics (Abbasi et al., 2016; Hurwitz et al., 2013). Indeed, inspecting the accuracy of users' generated data is challenging despite advances in semantic analytics and natural language processing (Abbasi et al., 2016). Therefore, we need to understand the factors that influence users' willingness to provide inaccurate information. We build on the privacy literature and introduce the construct of self-disclosure accuracy, which we define as users' willingness to provide accurate and complete personal information to social networking websites. Thus, we investigate:

RQ: How does familiarity with big data impact privacy concerns and disclosure outcomes in the context of social networking websites?

Information systems (IS) researchers have adopted various theories to study the construct of privacy concerns, its determinants, and its outcomes (for review, see Li, 2011, 2012). Smith, Dinev, and Xu (2011) review the privacy literature and provide an overarching APCO[1] model in which they suggest that privacy concerns mediate the relationship between several antecedents (e.g., privacy awareness) and behavioral outcomes (e.g., privacy-protective behaviors). Further, extant IS research has shown that protection motivation theory (PMT) and the theory of planned behavior (TPB) are well suited for studying behaviors related to fear appeals, such as privacy and security concerns (for references, see Table A1 and Table A2). In this study, we adapt Smith's et al. (2011) APCO model while integrating PMT and TPB to derive the theoretical links in our research model (see Figure 1).

Specifically, the research model considers familiarity with big data, perceived control, perceived vulnerability, and self-efficacy as antecedents of privacy concerns. The construct of privacy concerns mediates the effect of these antecedents on two outcomes: self-disclosure accuracy and self-disclosure concerns. The model also considers trust as a moderator of the relationship between privacy concerns and self-disclosure accuracy. We test the model using a factor-based structural equation modeling (SEM) technique. We also test other alternative models to examine the mediating role of privacy concerns, to control for demographic variables, and to investigate different roles of trust.

In this study, we examine users' privacy-related attitudes and outcomes. Research in the context of social networking websites continues to proliferate. However, a recent review of this literature suggests that we still lack research that focuses on interpersonal relationships, such as privacy attitudes (Cao et al., 2015). From a theoretical perspective, we shed light on how familiarity with big data plays a key role in determining privacy concerns in the context of social networking websites. In addition, we account for several other constructs drawn from PMT and TPB (i.e., perceived control, perceived vulnerability, self-efficacy, and trust) in one theoretical model. Furthermore, we provide practical implications in terms of the expected accuracy of social media big datasets. Big data adopters can be a social networking website or an external entity that uses social media's big data. Data accuracy is a critical practical issue in the big data analytics domain, which we discuss in Section 2.

This paper is organized as follows. First, we briefly discuss data analytics techniques and the issue of privacy and data accuracy. Next, we discuss theoretical backgrounds and the research model. Then, we present the methodology, data analysis, and results, followed by a discussion. Finally, we conclude with theoretical and practical implications and avenues for future research.

[1] APCO stands for "antecedents → privacy concerns → outcomes" (Smith et al., 2011)

2 Social Networking Websites and Big Data: The Issue of Privacy and Data Accuracy

Since their proliferation in the early 2000s, social networking websites have presented a fascinating milieu for both researchers and practitioners. Starting in 2004, IS researchers began investigating factors that influence social networking websites, developing algorithms and tools to analyze these websites, and examining the impact of social networking websites on individuals (Cao et al., 2015). Practitioners have been more interested in using this environment to optimize decision making and to gain a competitive advantage (Varadarajan & Soundarapandian, 2013). Before the official arrival of big data analytics in the early 2010s, practitioners and the service providers of social networking websites did actually operate some sort of big data analytics using different terminology, such as social media mining and text analytics (He et al., 2013; Hurwitz et al., 2013; Schmarzo, 2013). However, practitioners have encountered many obstacles in implementing social media mining and other types of analytics due to the challenging and intricate operations required to generate accurate and reliable insights (Varadarajan & Soundarapandian, 2013).

For example, performing text analytics on social networking websites involves searching for data (i.e., structured, semi-structured, or unstructured), extracting related data, and converting them into one structured format that one can use to generate insights (Hurwitz et al., 2013). This technology enables its adopters to gather and analyze different types of data about social networking website users. For instance, one can extract complete and specific information about users (e.g., name, address, phone number, location, affiliation, employment, feeling, and interests). One can them combine these data in a structured database that marketers and data brokers can use for commercial purposes. However, this process becomes very complicated when considering the deluge of data available on social networking websites, which leads to diminished accuracy due to limited technological capabilities and human involvement (Varadarajan & Soundarapandian, 2013). Furthermore, the issue of privacy has presented another dilemma that has imposed more restrictions on social media mining and, thus, amplified its limitations (Shim et al., 2015; Watson, 2014).

Today, when big data analytics dominate the IT industry, practitioners are much more powerful than before. Big data analytics not only encompass the preceding social media mining techniques but also provide much more advanced analytical techniques supported by highly scalable infrastructures (e.g., MapReduce, Hadoop, Big Table, Hive, and HBase) (Hurwitz et al., 2013; Schmarzo, 2013; Vera-Baquero, Colomo-Palacios, & Molloy, 2013; Watson, 2014). These advanced techniques have enabled practitioners to tackle the issue of high volume data. As a result, harnessing social media data has boomed, and social media has become among the top five sources used by big data adopters (IBM, 2012; Kart, Heudecker, & Buytendijk, 2013). Big data advocates promote the promising benefits of big data applications to both users and practitioners (Hurwitz et al., 2013; Manyika et al., 2011; Russom, 2011). In contrast, many scholars and business leaders have questioned big data applications and predicted that big data will exacerbate already-complicated privacy issues (Bertolucci, 2013; Boyd & Crawford, 2012; Breznitz et al., 2011; Craig & Ludloff, 2011; Shim et al., 2015; Watson, 2014). The latter maintain that big data analytics and their unprecedented implications, including collecting, storing, analyzing, sharing, and monetizing personal data, will most likely change our understanding of information privacy in the social media world and increase privacy concerns overall.

Big data as a technology provides powerful capabilities in terms of capturing, storing, managing, and analyzing data of high volume (i.e., how much data), variety (i.e., various types of data), and velocity (i.e., how fast data are processed)—the so-called “three Vs” (Beyer & Laney, 2012; Hurwitz et al., 2013; Manyika et al., 2011). These three characteristics, or “Vs”, represent the main merits of big data analytics. However, there are other “Vs” that, if not reinforced, can significantly undermine the capabilities that the first three provide. Other “Vs” include but may not be limited to: veracity, validity, volatility, viability, and value. In this study, we focus on veracity and validity. These two “Vs” are highly interrelated because both reflect data’s accuracy (Hurwitz et al., 2013). Validity refers to the accuracy of the data fed into big data tools and veracity refers to the accuracy of insights generated by big data analytics. We maintain that user-generated data partially affect both veracity and validity. In other words, when users provide inaccurate (accurate) personal data, they weaken (strengthen) both veracity and validity.

Because social media as a big data source largely depends on user-generated or self-disclosed data, studying users of social networking websites is one direct way to identify factors that affect the accuracy of personal information. According to Hurwitz et al. (2013), decision makers rely on analysts who should be

extra vigilant regarding validity when it comes to moving big data from exploration to action. Although the data-validation process includes different phases in practice, investigations of one of the main root sources of inaccurate data (i.e., users) and attempts to minimize those inaccuracies could be more efficient. By providing insights about the factors that contribute to weakening the accuracy of self-disclosed data, practitioners can work on mitigating them through policy changes and other IT solutions. Such actions would increase the accuracy of data fed into big data tools (i.e., validity) and, consequently, enhance the accuracy of insights generated by big data analytics (i.e., veracity).

3 Research Model, Theoretical Backgrounds, and Related Work

Smith's et al. (2011) APCO model suggests that the construct of privacy concerns will likely mediate the relationships between a set of antecedents (e.g., privacy experience, privacy awareness, and personality) and behavioral outcomes (e.g., self-disclosure, risks, and regulation). They indicate that little research has tested the relationship between privacy awareness and privacy concerns. Our main theoretical contribution is to shed light on this relationship by studying how awareness of the concept and practices of big data impacts privacy concerns.

As Figure 1 shows, privacy concerns is the focal construct that serves as a mediator between a set of antecedents (i.e., familiarity with big data, perceived control, perceived vulnerability, and self-efficacy) and two outcomes (i.e., self-disclosure accuracy and self-disclosure concerns). We model trust as a moderator of the relationship between privacy concerns and self-disclosure accuracy. Next, we discuss PMT and TPB and relevant empirical work to theorize these relationships.

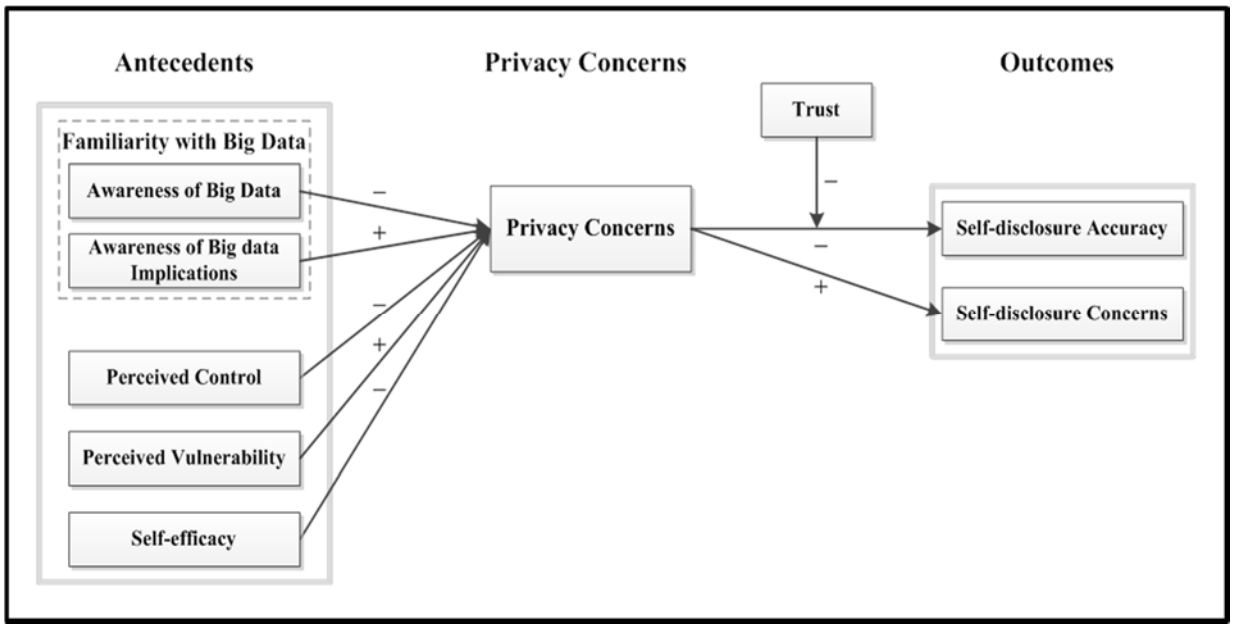

Figure 1. Antecedents → Privacy Concerns → Outcomes (APCO) Model

3.1 Theoretical Backgrounds

IS researchers have drawn on various theories to study fear appeals. Examples of fear appeals are situations that involve potential loss of information privacy or security. In this regard, two theories stand out as a promising lens: protection motivation theory (PMT) and the theory of planned behavior (TPB); the latter extends the theory of reasoned action (TRA) (Ajzen, 1991; Ajzen & Fishbein, 1980; Rogers, 1975; Maddux & Rogers, 1983). Tables A1 and A2 summarizes relevant research that builds on these two theories.

3.1.1 Protection Motivation Theory

As Floyd, Prentice-Dunn, and Rogers (2000, p. 409) state: "The protection motivation concept involves any threat for which there is an effective recommended response that can be carried out by the individual". PMT's core assumptions suggest that individuals, when confronted with a threatening

situation, go through two main processes: threat appraisal and coping appraisal. In the threat-appraisal process, individuals weigh the perceived severity and vulnerability of the threat against the rewards. When the perceived severity and perceived vulnerability exceed the expected rewards, individuals will seek protection by following a coping-appraisal process. For individuals to pursue the coping process and engage in the protective behavior, their self-efficacy and response efficacy must exceed the response costs. According to PMT, individuals differ in terms of evaluating the sensitivity and vulnerability of a certain situation and how they react to threatening situations.

Many IS researchers have embraced PMT to study users' protective behaviors by either adopting its full extent (e.g., Boss, Galletta, Lowry, Moody, & Polak, 2015) or by using some derivations of PMT (e.g., Crossler & Bélanger 2014). After extensively reviewing PMT and its application in the IS security research, Boss et al. (2015) tested the core and full PMT nomology in two different studies in the contexts of data backups (study 1) and anti-malware software (study 2). In both studies, the results fully supported PMT assumptions when fear-appeal manipulation was high, which suggests that PMT is a useful and suitable theoretical foundation for studying protective intentions and behaviors. Crossler and Bélanger (2014) show that PMT is effective in explaining several individual security practices.

In the IS privacy literature, however, most studies (if not all) have used derivations of PMT to contextualize the construct of privacy concerns as a fear appeal or threat. For instance, Junglas, Johnson, and Spitzmüller (2008) used PMT to theorize relationships between personality and concern for privacy. They argue that "since personality traits are resistant to transformation, it means that concerns about threats are explainable, to at least some extent, by individual's personality traits" (p. 391). Kuo, Ma, and Alexander (2014) adapted PMT in a similar manner and tested the relationships between privacy concern dimensions and privacy-protective behaviors, such as intent to falsify personal health information. Dinev and Hart (2004) found that perceived vulnerability is significantly associated with perceived privacy concerns but did not test how other PMT constructs affect perceived privacy concerns. Other researchers have focused on ethics-related factors, such as perceived moral judgment of an employer's requesting access to social media information, and how they influence privacy protection intentions (Drake et al., 2016). In the communication literature, some scholars have tested the relationships between all PMT constructs except response costs in the context of social networking websites but found only partial support for PMT (Mohamed & Ahmad, 2012). We could not find an IS privacy study that adhered to the full nomology of PMT or at least encompassed most of its constructs. Indeed, a full test of PMT in the privacy literature would further our understanding of privacy-protective behaviors. However, previous studies have mainly focused on other matters such as scale development (Dinev & Hart, 2004). In a similar vein, we adapt PMT to contextualize privacy concerns as a threat source that triggers privacy-protective behaviors. A full test of PMT is outside the scope of our study. Nevertheless, we test how two PMT constructs (i.e., perceived vulnerability and self-efficacy) influence privacy concerns.

3.1.2 The Theory of Planned Behavior

TPB suggests that "attitude toward the behavior, subjective norms, with respect to the behavior, and perceived control over the behavior are usually found to predict behavioral intentions with a high degree of accuracy" (Ajzen, 1991, p. 206). Researchers from different social science domains have used TPB extensively, which suggests that it is a useful theoretical lens to explain humans' behaviors (Ajzen, 2011). For instance, researchers have used TPB, or its former version, TRA, to explain adoption of ubiquitous commerce (Sheng, Nah, & Siau, 2008), e-commerce (Liu, Marchewka, Lu, & Yu, 2005), mobile commerce (Mishra, 2014), RFID technologies (Cazier, Jensen, & Dave, 2008), and instant messaging (IM) technologies (Lowry, Cao, & Everard, 2011).

Researchers in the IS privacy literature have also used TPB to explain users' intentions to disclose personal information. Generally, researchers have contextualized the construct of privacy concerns as a salient dispositional belief that influences behavioral intentions to disclose personal information (Bansal, Zahedi, & Gefen, 2016). In other words, privacy concerns represent the attitude toward the behavior in TPB jargon. Several studies have shown that privacy concerns and perceived control are significant predictors of intentions to disclose personal information (see Tables A2 and A3). However, they have not shown support for the influence of subjective norms in the context of privacy disclosure (Li & Slee, 2014; Xu, Michael, & Chen, 2013). A preliminary review of the IS privacy literature suggests that the majority of prominent IS privacy studies have relied on the assumption of TPB ("intention → actual disclosure behavior") (Alashoor, Lambert, & Farivar, 2016). Although intention is not a perfect predictor of actual behavior, several recent privacy and security studies have found a significant relationship between

intention and actual behavior (Boss et al., 2015; Keith, Thompson, Hale, & Lowry, 2013). Similarly, Malhotra, Kim, and Agarwal (2004) suggest that "intention to release personal information serves as a good proxy for whether one actually reveals personal information at the request of an online marketer" (p. 342). Based on our review of this literature, we derive two constructs from TPB: perceived control and privacy concerns.

To summarize, we integrate PMT and TPB to develop a research model that predicts users' willingness to provide accurate personal information (i.e., self-disclosure accuracy) in the context of social networking websites. Based on Smith's et al. (2011) APCO model, we position the construct of privacy concerns as a mediator. Specifically, we propose that privacy concerns will fully mediate the impact of the constructs drawn from PMT and TPB and the newly presented construct (i.e., familiarity with big data) on self-disclosure accuracy. We also propose that privacy concerns will have an effect on self-disclosure concerns. In Sections 3.3 to 3.6, we discuss these hypotheses in more detail.

3.2 Privacy Concerns

Researchers have defined the construct of privacy concerns in many different ways (Hong & Thong, 2013; Malhotra et al., 2004; Smith, Milberg, & Burke, 1996; Steinbart, Keith, & Babb, 2017; Stewart & Segars, 2002); for a review, see Bélanger and Crossler (2011), Li (2011), and Smith et al. (2011). Nevertheless, the definition of privacy concerns in most empirical studies reflects individuals' perceptions of the loss of privacy or the limited level of privacy protection in online contexts (Smith et al., 1996). In various online contexts, IS privacy researchers have shown that a number of antecedents predict the construct of privacy concerns, which, in turn, predicts privacy-protective behaviors and self-disclosure outcomes. We define privacy concerns as "concerns about opportunistic behavior related to the personal information submitted over (social networking websites) by the respondent in particular" (Dinev & Hart, 2006, p. 64). Table A3 summarizes related constructs tested in privacy research in the context of social networking websites and Table A4 summarizes the definitions of constructs we tested in this study.

3.3 Self-disclosure Concerns and Self-disclosure Accuracy

The nature of social media communications necessitates and motivates self-disclosure behaviors. Social networking websites present valuable socializing opportunities and enable their users to interact, communicate, and share information with each other (Acquisti & Gross, 2006). Without information disclosure, users can barely perceive the benefits of using social media. As a result, users tend to reveal personal information, (e.g., name, age, city, email, personal photos, life experiences, feelings, and other types of personal information) in order to become engaged in a social networking website. Yet, research has widely shown that concerns for privacy negatively influence self-disclosure outcomes. In other words, concerned users are less willing to reveal personal information (Dinev & Hart, 2006; Keith, Babb, Lowry, Furner, & Abdullat, 2015; Malhotra et al., 2004; Sheehan & Hoy, 1999; Smith et al., 1996). Several studies support this relationship in the context of social networking websites (Baruh & Cemalcilar, 2014; Krasnova, Spiekermann, Koroleva, & Hildebrand, 2010; Tow, Dell, & Venable, 2010; Xu et al., 2013).

Users, however, do not always follow a rational privacy model. They are likely to reveal personal information, sometimes highly sensitive, as they become cognitively absorbed in the social networking activity (Alashoor & Baskerville, 2015). In such cases, users do not adhere to their privacy beliefs and are likely to behave contradictorily to them. However, when users rethink their disclosure behaviors and the amount of personal information they have shared, they start to feel concerned about their personal information that is available on the network and accessible to many others. We refer to such feelings as self-disclosure concerns. This construct reflects users' concerns about the extent of personal information they have revealed to social networking websites. Self-disclosure concerns differ from privacy concerns in the sense that the latter represents concerns about organizational practices related to the use of personal information (Dinev & Hart, 2006; Smith et al., 1996). Self-disclosure concerns, on the other hand, represent concerns about the disclosing activity per se. We predict a positive relationship between these two constructs such that users with high privacy concerns are likely to be also highly concerned about their self-disclosure behaviors on social networking websites.

H1: Privacy concerns are positively associated with self-disclosure concerns.

Self-disclosure accuracy refers to the users' willingness to provide accurate and complete personal information to social networking websites. Fundamentally, disclosing personal information to social networking websites means that these websites (and possibly other parties) will use users' personal

information for big data purposes since prominent social networking websites have already adopted big data analytics (Tan, Blake, Saleh, & Dustdar, 2013). Accordingly, our conceptualization of the self-disclosure accuracy construct encompasses the two interchangeable contexts: social networking websites and big data. Willingness to falsify personal information and refusals to give out personal information appropriately reflect the potential level of accuracy. Thus, we leverage related research that has investigated these two elements.

A handful of empirical studies have examined the relationship between privacy concerns and willingness to falsify personal information. Findings from this literature are also mixed. For instance, Xie, Teo, and Wan (2006) found that Internet users tend to falsify their personal information when online retailers ask them for it. Users falsify personal information because it is much more convenient than reporting a privacy dispute to a third party privacy organization (Lwin & Williams, 2003). Other studies also support this relationship in different contexts, such as work settings, chat rooms, mobile apps, and social networking websites (e.g., Jiang, Heng, & Choi, 2013; Keith et al., 2013; Keith, Maynes, Lowry, & Babb, 2014; Mohamed, 2010; Posey, Bennett, Roberts, & Lowry, 2011). In contrast, other studies have failed to provide support for this relationship among Internet and social networking website users (e.g., Acquisti & Gross, 2006; Drake et al., 2016; Sheehan & Hoy, 1999; Son & Kim, 2008; Youn, 2009). With regard to willingness to provide complete personal information, the literature provides strong support for a negative relationship. Specifically, concerned users are likely to avoid giving personal information (Dinev & Hart, 2006; Drake et al., 2016; Keith et al., 2015; Keith, Babb, Furner, Abdullat, & Lowry, 2016; Krasnova et al., 2010; Lowry et al., 2011; Posey, Lowry, Roberts, & Ellis, 2010; Sharma & Crossler, 2014; Sheehan & Hoy, 1999; Son & Kim, 2008; Tow et al., 2010; Xu et al., 2013; Youn, 2009).

Research has found social networking website users provide inaccurate, unreliable, and incomplete information to protect themselves from potential frauds and identity thefts or due to their privacy concerns. As Tow et al. (2010, p. 133) report one Facebook user's saying: "Identity theft would be one of my concerns and that is why I never publish more full details nor do I publish everything correctly". In some contexts, such as online retailer websites, users might not be able to transact successfully unless they provide accurate and complete information. However, "self-disclosure is one of the most supported boundary mechanisms in SNS interfaces and is often characterized as the 'privacy settings' of one's user profile" (Wisniewski, Islam, Lipford, & Wilson, 2016, pp. 239). In other words, social networking website users may protect themselves from potential privacy and security threats by providing false and incomplete information. Based on this literature, we predict a negative relationship between privacy concerns and self-disclosure accuracy.

H2: Privacy concerns are negatively associated with self-disclosure accuracy.

3.4 Trust

Trust is a strong predictor of online purchasing intentions and behaviors (Ba & Pavlou, 2002; Bélanger, Hiller, & Smith, 2002; Gefen, Rao, & Tractinsky, 2003; Pavlou & Fygenson, 2006). Online companies can build trust with consumers through logos, branding, and website quality (Lowry, Roberts, & Caine, 2005; Lowry, Vance, Moody, Beckman, & Read, 2008; Lowry, Wilson, & Haig, 2014). Further, research shows that privacy assurance mechanisms play a significant role in online disclosure and purchasing contexts (Bansal, Zahedi, & Gefen, 2015; Lowry et al., 2012; Posey et al., 2010; Wu, Huang, Yen, & Popova, 2012). Privacy researchers argue that trust is an essential element in any privacy-related context because disclosure behaviors involve some degree of risk (Dinev & Hart, 2006). Empirical evidence shows that trust in Internet websites is positively associated with individuals' willingness to provide personal information to transact online (Dinev & Hart, 2006). The effect size of trust is about twice the negative effect of privacy concerns, which suggests that trust may even revoke privacy concerns when it comes to transacting online (Dinev & Hart, 2006). Further, trust in social networking websites is negatively associated with privacy concerns and positively associated with self-disclosure outcomes (Krasnova et al., 2010; Krasnova, Veltri, & Günther, 2012). Still, other factors, such as the strength of ties in a social network, could influence the users' overall trust level in a network (Bapna, Gupta, Rice, & Sundararajan, forthcoming).

Researchers have modeled trust in different ways in the privacy literature (Smith et al., 2011). Studies have treated trust as a determinant of privacy concerns (Krasnova et al., 2010; Li, Gupta, Zhang, & Sarathy, 2014; Taddei & Contena, 2013), a determinant of disclosure outcomes (Bansal et al., 2015; Krasnova et al., 2012), a mediator between privacy concerns and disclosure outcomes (Bansal et al., 2016; Liu et al., 2005; Malhotra et al., 2004), and a moderator of the relationship between privacy

concerns and disclosure outcomes (Bansal, Zahedi, & Gefen, 2008; Taddei & Contena, 2013). The fact that privacy studies have modeled trust differently suggests that examining the impact of trust in privacy disclosure contexts is complicated.

With respect to the differences in this literature, we believe that it is theoretically insightful to treat trust as a moderator between privacy concerns and disclosure outcomes. Indeed, Taddei and Contena (2013), who tested how trust operates in different models, have previously pointed this point out. In addition, the continuing concerns for privacy and the high level of self-disclosure behaviors among social networking website users represent a privacy paradox (Smith et al., 2011). Without considering affective factors that can explain this paradox, the negative relationship between privacy concerns and disclosure outcomes would not provide insightful results since users tend to disclose much personal information in practice. Accordingly, considering trust as a moderator can enrich our understanding of the negative relationship between privacy concerns and disclosure outcomes.

As we hypothesize above, users who have privacy concerns would likely be less willing to give out personal information and more willing to falsify their information and, hence, be less likely to provide accurate information. We define trust as the extent to which users are confident that social networking websites will competently, reliably, and safely handle their personal information (Dinev & Hart, 2006). Thus, high levels of trust can, theoretically and statistically, attenuate the negative effect of privacy concerns. Modeling trust as a moderator would provide important information as to when or at what levels of trust privacy concerns negatively impact the accuracy of information that users give. We predict that the strength of the negative relationship between privacy concerns and self-disclosure accuracy will weaken with high levels of trust.

> **H3:** Trust moderates the negative relationship between privacy concerns and self-disclosure accuracy such that the relationship is weaker (stronger) when trust is high (low).

3.5 Familiarity with Big Data

According to Craig and Ludloff (2011), it is not wrong or immoral for businesses to collect, store, analyze, or even share social media users' personal information and sentiments because they do so to improve business and satisfy customers. However, these pursuits are susceptible to misconduct and human error. Furthermore, intruders and malicious individuals can use several free analytic tools for social media mining (He et al., 2013) to harm users. Various Web-based commercial analytic tools exist to provide such services for free or a modest amount of money. Currently, the leading social networking websites may own the biggest datasets in our digital world (e.g., Facebook's Presto and Twitter's Storm). By adopting big data analytics, social networking websites can collect and store exabytes of data about their users (Shim et al., 2015; Tan et al., 2013). Then, they can analyze and share this high volume of data for various reasons with or without users' being fully aware of how their data is being used (Craig & Ludloff, 2011). Nowadays, one can easily track the location of users' posts and tweets. The truth is:

> *There is an unprecedented aggregation of data about each one of us available in digital format. This makes it easy for organizations of all sizes, as well as governmental agencies, to find information about any individual as well as use analytic models to predict future behavior.* (Craig & Ludloff, 2011, pp. 5)

Therefore, the emergence of big data and its unprecedented implications will most likely change our understanding of information privacy in the social media world. Many scholars and business leaders have already raised the issue of big data privacy (Bertolucci, 2013; Boyd & Crawford, 2012; Breznitz et al., 2011; Craig & Ludloff, 2011; Shim et al., 2015; Watson, 2014). Clemons, Wilson, and Jin (2014) show that individuals have little understanding and concern about big data and its implications. Researchers have also predicted that, as individuals become aware of big data, they are likely to express more concerns for privacy (Clemons et al., 2014; Watson, 2014). However, empirical evidence shows how familiarity with big data impacts individuals' privacy concerns. As such, we explore this relationship in this study.

Familiarity with big data refers to individuals' awareness of the term big data, including its values to them and to businesses, and the four major implications of big data analytics (i.e., collection, storing, processing, and sharing). Many individuals are already aware that online companies already exploit their personal information, a perception that explains their concerns for privacy and disclosure behaviors (Choi & Land, 2016; Krasnova, Günther, Spiekermann, & Koroleva, 2009; Malhotra et al., 2004; Turow & Hennessy, 2007). However, users might also be aware that big data may impart some future benefits to them, which suggests two opposing foci of the familiarity with big data construct referred to as 1)

awareness of big data and 2) awareness of big data implications. In particular, individuals who are aware of big data and its benefits are less likely to be concerned because they perceive the fact that one has to give up some personal privacy in order to realize the benefits of today's technology. This prediction is in line with the negative association that research has found between Internet literacy and privacy concerns (Dinev & Hart, 2005). Krasnova, Veltri, and Garah (2014) also found partial support for a negative association between awareness of information handling by social networking websites and privacy concerns. In contrast, being aware of the ubiquitous uses of personal data at the same time can lead to concerns about personal privacy and disclosing behaviors (Krasnova et al., 2009). Recent polls support these contradictory personal beliefs among U.S. citizens with regard to governmental surveillance: they have found that people generally approve of governmental surveillance for its security benefits but are still concerned about the privacy of personal communications (Gao, 2015). Thus, we propose that awareness of big data and its benefits will contribute to reducing privacy concerns. However, we propose that awareness of big data implications will contribute to increasing privacy concerns.

H4: Awareness of big data is negatively associated with privacy concerns.

H5: Awareness of big data implications is positively associated with privacy concerns.

3.6 Perceived Control, Perceived Vulnerability, and Self-efficacy

The ability to possess control over personal information is the user's right, and social networking websites should give this right to their users (Dinev & Hart, 2004). The ability to control information is one of the key components of information privacy, and IS privacy researchers have used the control aspect to define privacy concerns (Hong & Thong 2013; Malhotra et al. 2004; Smith et al. 1996; Stewart & Segars 2002). Perceived control refers to the one's perceived ability to control personal data submitted to social networking websites (Dinev & Hart, 2004). Consistent with TPB's predictions (Ajzen, 1991), the privacy literature shows a strong support for the mitigating effect of perceived control on privacy concerns. Research has established this support in various contexts, including social networking websites (Cavusoglu, Phan, Cavusoglu, & Airoldi, 2016; Krasnova et al., 2010; Krasnova et al., 2014; Wisniewski et al., 2016; Xu et al., 2013). In line with this literature, we predict a negative association between perceived control and privacy concerns.

H6: Perceived control is negatively associated with privacy concerns.

Perceived vulnerability, a construct derived from PMT (Rogers, 1975), refers to the perceived potential risks associated with revealing personal information to social networking websites. Although few studies have tested the relationship between perceived vulnerability and privacy concerns (see Table A3), research has found that perceived vulnerability is a significant determinant of privacy concerns in the Internet context (Dinev & Hart, 2004; Youn, 2009) and social networking websites (Mohamed & Ahmad, 2012). In the security domain, Boss et al. (2015) also found support for this relationship such that both users' fear appeals and security protective intentions increase when they perceive the situation as vulnerable. Thus, we predict a positive association between perceived vulnerability and privacy concerns.

H7: Perceived vulnerability is positively associated with privacy concerns.

Self-efficacy, also derived from PMT (Rogers, 1975), refers to users' confidence and competence to cope with potential privacy threats on social networking websites. For instance, users who can use the privacy settings that social networking websites provide are likely to employ them as a way to protect their privacy (Mohamed & Ahmad, 2012). Similarly, users with high self-efficacy in using protective software programs tend to use them for protective purposes (Boss et al., 2015; Crossler & Bélanger, 2009, 2014). Keith et al. (2015) have also found that mobile-computing self-efficacy is negatively associated with perceived privacy risk, which, in turn, impacts disclosure outcomes. The few studies that have studied the effect of self-efficacy on privacy concerns have report mixed results. Specifically, Mohamed and Ahmad (2012) proposed and found support for a positive relationship between self-efficacy and privacy concerns. Youn (2009) also proposed a positive relationship but found no support. Further, Yao, Rice, and Wallis (2007) proposed and found support for a negative relationship between self-efficacy and privacy concerns. Both Mohamed and Ahmad (2012) and Youn (2009) used PMT predictions, but Yao et al. (2007) adopted social cognitive theory (SCT) (Bandura, 1988). From both theoretical perspectives, individuals who perceive that they can handle a threatening situation will likely tend to adopt protective behaviors, which will lead to a much lower level of anxiety. We rely on this prediction and propose that users who have high self-efficacy in terms of protecting their personal information on social networking websites will likely experience less concerns for privacy.

H8: Self-efficacy is negatively associated with privacy concerns.

4 Method and Sample Characteristics

We used a cross-sectional survey design to collect data that measured items that reflect the constructs we discuss in Section 3. We also measured several demographic variables (Table 1). The targeted sample included undergraduate and graduate students at a public university in the northeast United States. We justify our using students as subjects based on the popularity of using social networking websites among college students. In addition, because the survey measured some items that involved idiosyncratic terms (e.g., big data), using students ensured our subjects could adequately understand the items we used. For instance, most students, especially those in the business school, took business courses that taught big data.

Table 1. Demographics Descriptive Statistics

	N	Percent		N	Percent
Gender			**Employment**		
Male	159	61.4	Full-time	57	22.0
Female	84	32.4	Part-time	107	41.3
Did not report	16	6.2	Not currently employed	64	24.7
Age			Retired	2	0.8
Under 18	2	0.8	Did not report	29	11.2
18 – 21	116	44.8	**Marital status**		
22 – 25	77	29.7	Never married	214	82.6
26 – 29	27	10.4	Married	24	9.3
30 – 35	15	5.8	Divorced	5	1.9
36 and above	8	3.1	Did not report	16	6.2
Did not Report	14	5.4	**Race 1**		
Education			Not Hispanic or Latino	211	81.5
Freshman	39	15.1	Hispanic or Latino	20	7.7
Sophomore	50	19.3	Did not report	28	10.8
Junior	42	16.2	**Race 2**		
Senior	60	23.2	White	138	53.3
Master's	51	19.7	Black or African American	31	12.0
PhD	3	1.2	Asian	50	19.3
Did not report	14	5.4	Native Hawaiian or other Pacific Islander	2	0.8
Social networks use			Did not report	38	14.7
Never	17	6.6			
< 1 hour per day	82	31.7			
1 to 2 hours per day	68	26.3			
2 to 3 hours per day	33	12.7			
> 3 hours per day	47	18.1			
Did not report	12	4.6			

We conducted a pilot study to ensure the clarity of the items; we dropped none of the items. The institutional review board (IRB) approved the questionnaire prior to our administering it, and we provided no incentives for participating in this study. We distributed a paper-based questionnaire to 275 students during their class period. We explained the study's purpose to students before they participated. Fifteen students did not participate, and we removed one response from the final dataset because it was clearly not a reliable one. Some of the returned surveys included several missing values. The final sample used

to conduct the main analysis comprised 208 surveys[2]. Table 1 presents the descriptive statistics for demographic variables.

5 Measurements, Data Analysis, and Results

5.1 Preliminary Analysis

We adapted the measurements from previous research whenever possible. We developed the items for awareness of big data. Table B1 presents the measurements we used and their scales and sources. We conducted a factor-based SEM and confirmatory factor analysis using the lavaan package in R. The preliminary analyses showed that the model demonstrated poor fit. Accordingly, we examined the correlation matrix using the proportionality principles and followed the re-specification rules that Kline (2011) recommends. According to the preliminary investigations of the correlations and loadings of the observed variables, three items (PeC3, SDC2, and SDA3) showed low loadings on the constructs they were supposed to reflect. Therefore, we omitted these indicators. In addition, the model contained an interaction term, and modeling such a term in the lavaan package requires one to use a consistent number of indicators for each of the interacting factors. Because we used only three indicators to measure trust and four indicators to measure privacy concerns, we had to eliminate one item that loaded the least on the privacy concerns construct (PrC2) in order to model the moderating effect appropriately (three indicators per each factor). Finally, we excluded responses from subjects who indicated that they never use social networking websites. The supplementary documents present the covariance/correlation matrices of the observed/unobserved variables. In Sections 5.2 to 5.4, we discuss the validation of the measurement model and the results from four structural models.

5.2 Measurement Validation and Confirmatory Factor Analysis

Table 2 presents the standard deviations of the unobserved variables, their correlations, and reliability measurements. The Cronbach's alpha reliabilities were all above the .7 threshold and they ranged from .746 (self-disclosure accuracy) to .967 (awareness of big data). All average variance extracted (AVE) scores were above .5 and ranged from .557 (perceived vulnerability) to .881 (awareness of big data). All AVE scores were higher than any correlation with other latent variables. One exception was that the AVE score for self-disclosure concerns was just below its correlation with privacy concerns. To assess this issue, we examined the correlation matrix of the observed variables and used proportionality principles (Kline, 2011). It appeared that the observed variables for self-disclosure concerns and privacy concerns did not tap equally onto one latent variable and the ratios differed to a large extent. According to these results, we concluded that both convergent and discriminant validities were satisfied.

Table 2. Correlations and Measurement Reliabilities of Unobserved Variables

	Means	S.D.	Alpha	SDA	SDC	PrC	T	ABD	ABDI	PeC	PV	SE
SDA	3.516	1.159	0.746	**0.589**								
SDC	5.407	0.723	0.826	-0.274	**0.618**							
PrC	4.091	0.622	0.892	-0.398	0.666	**0.742**						
T	2.395	0.839	0.830	0.088	-0.107	-0.161	**0.622**					
ABD	2.976	1.296	0.967	0.013	-0.035	-0.053	-0.140	**0.881**				
ABDI	4.388	0.644	0.937	-0.099	0.175	0.263	-0.072	0.199	**0.792**			
PeC	3.886	0.961	0.787	0.035	-0.053	-0.079	0.054	-0.058	0.103	**0.590**		
PV	4.215	0.41	0.856	-0.200	0.337	0.506	-0.250	0.045	0.247	0.107	**0.557**	
SE	3.331	1.062	0.775	0.118	-0.155	-0.233	0.355	-0.185	-0.022	0.271	-0.193	**0.669**

* Items on the diagonal represent AVE scores
***** SDA: self-disclosure accuracy, SDC: self-disclosure concerns, PrC: privacy concerns, T: trust, ABD: awareness of big data, ABDI: awareness of big data implications, PeC: perceived control, PV: perceived vulnerability, SE: self-efficacy.

[2] The main analysis deals with the research model that we depict in Figure 1. We also conducted other analyses (see the following sections) to control for some demographic variables. Because some returned surveys had other missing values in the demographic measures, the sample size for the model that controlled for demographics comprised 182 surveys.

While we focused predominantly on testing the research model we depict in Figure 1 in this study, we also tested other alternative models to1) test the mediation effect of privacy concerns, 2) control for some demographic variables, and 3) investigate different roles of trust. Table B2 presents the loadings of the unobserved variables for each model based on a confirmatory factor analysis approach. Model 1 represents the main model (Figure 1). Model 2 represents a model in which we included direct paths from the antecedents of privacy concerns to self-disclosure accuracy and self-disclosure concerns. Model 3 represents a model in which we controlled for four demographic variables by adding four paths to self-disclosure accuracy and self-disclosure concerns[3]. Model 4 represents a model in which we treated trust as a predictor of privacy concerns, self-disclosure accuracy, and self-disclosure concerns. All models showed significant loadings ($p < .000$), and, hence, we proceeded to evaluate the models' fit and path coefficient results.

5.3 Fit of Structural Models

Table 3 presents the results from a series of SEM analyses along with their fit indices. We used a maximum likelihood robust (MLR) estimator to correct for non-normality violations. Although all models indicated overall misfit based on chi-squares, one should not use only this statistic to judge the overall fit of a model. Rather, one should use other approximate fit indices (e.g., GFI, AGFI, CFI, RMSEA, and SRMR*)* to assess the overall fit (Hu & Bentler, 1999; Kline, 2011). We expected inflated chi-squares because of the large number of indicators involved in this study, which subsequently enlarged degrees of freedom. As a result, chi-squares would not have helped us to assess the overall fit, especially when the assumptions did not hold. Accordingly, to assess the overall fit of the tested models, we followed the combination rule (Hu & Bentler, 1999), which suggests that a good approximate fit should show that SRMR is less than or equal to 0.08 and that either CFI is above or equal to 0.95 or RMSEA is less than or equal to 0.06. All models satisfied this combination rule; hence, we concluded that all models exhibited a good approximate overall fit.

5.4 Hypothesis Tests

Model 1 tested the main research model that we depict in Figure 1. The approximate fit indices indicated that this model demonstrated good fit (GFI = 0.981, AGFI = 0.975, CFI = 0.944, RMSEA = 0.050, SRMR = 0.053). With regard to Model 1's path coefficients, PrC was significantly associated with all predictors at the 0.05 significance level (ABD → PrC estimate = -0.082; se = 0.037; ABDI → PrC estimate = 0.202; se = 0.083; PeC → PrC estimate = -0.093; se = 0.045; PV → PrC estimate = 0.662; se = 0.169). One exception was the path from self-efficacy, which was significant only at the 0.1 level (SE → PrC estimate = -0.095; se = 0.050). The two paths from PrC to self-disclosure concerns and self-disclosure accuracy were also significant (PrC → SDC estimate = 0.924; se = 0.122; PrC → SDA estimate = -0.673; se = 0.169). However, the results indicate that the moderating effect of trust was not significant (T x PrC estimate = 0.155; se = 0.190). Consequently, these results provide support for all proposed hypotheses except H3.

Model 2 tested the same model while incorporating direct paths from ABD, ABDI, PeC, PV, and SE to the outcome constructs (i.e., SDC and SDA). The approximate fit indices indicated that this model demonstrated good fit (GFI = 0.981, AGFI = 0.975, CFI = 0.943, RMSEA = 0.051, SRMR = 0.050). In terms of significance, the results did not differ from those observed in Model 1. While the coefficient estimates differed slightly from those in Model 1, the substantive interpretations of the findings would still be the same. More importantly, because we used this model to investigate the mediation role of privacy concerns, the results support the hypothesis that privacy concern fully mediates the relationships between its antecedents and outcomes (Smith et al., 2011). As Table 3 shows (Model 2), none of the direct path coefficients was significant. We also conducted a chi-square difference test to test whether Model 1 and Model 2 exhibited a difference in terms of model fit. The result from this test indicated no difference ($\chi_{(ML)}$ diff = 4.99; p = 0.896), which provides support for the full-mediation effect of privacy concerns.

[3] In Model 3, we added no direct paths from the antecedents of privacy concerns. We recoded education and ethnicity into binary variables because of the observed low percentages in many categories: education (graduates = 1; undergraduates = 0) and ethnicity (White = 1; non-White = 0). We treated age as a continuous variable: we coded females as 1 and males as 0.

Table 3. SEM Results

	Model 1	Model 2	Model 3	Model 4
Privacy concerns (PrC) R^2	32.7%	31.4%	33.1%	32.1%
ABD	-0.082 (0.037)*	-0.079 (0.038)*	-0.094 (0.039)*	-0.083 (0.038)*
ABDI	0.202 (0.083)*	0.198 (0.084)*	0.215 (0.098)*	0.201 (0.082)*
PeC	-0.093 (0.045)*	-0.094 (0.045)*	-0.103 (0.046)*	-0.099 (0.045)*
PV	0.662 (0.169)***	0.645 (0.171)***	0.742 (0.201)***	0.651 (0.166)***
SE	-0.095 (0.050)†	-0.092 (0.049)†	-0.069 (0.052)	-0.070 (0.053)
T	-	-	-	-0.034 (0.072)
Self-disclosure concerns (SDC) R^2	44.3%	45.5%	46.1%	45.4%
PrC	0.924 (0.122)***	0.784 (0.136)***	0.875 (0.131)***	0.886 (0.119)***
ABD	-	-0.065 (0.056)	-	-
ABDI	-	0.047 (0.124)	-	-
PeC	-	0.013 (0.092)	-	-
PV	-	0.259 (0.202)	-	-
SE	-	-0.084 (0.094)	-	-
Female	-	-	0.084 (0.145)	-
Age	-	-	0.001 (0.012)	-
Graduate	-	-	0.040 (0.175)	-
White	-	-	-0.048 (0.134)	-
T	-	-	-	-0.157 (0.088)†
Self-disclosure accuracy (SDA) R^2	16.3%	16.8%	20.9%	16.1%
PrC	-0.673 (0.169)***	-0.685 (0.181)***	-0.664 (0.187)***	-0.674 (0.165)***
T	0.008 (0.132)	-0.004 (0.138)	0.004 (0.141)	0.049 (0.121)
T x PrC	0.155 (0.190)	0.118 (0.198)	0.219 (0.205)	-
ABD	-	-0.039 (0.087)	-	-
ABDI	-	-0.026 (0.138)	-	-
PeC	-	-0.024 (0.111)	-	-
PV	-	0.041 (0.229)	-	-
SE	-	0.042 (0.128)	-	-
Female	-	-	0.103 (0.208)	-
Age	-	-	0.041 (0.021)*	-
Graduate	-	-	-0.356 (0.293)	-
White	-	-	0.295 (0.196)	-
Fit indices				
χ (ML)	705.170***	699.431***	932.744***	563.807***
χ (MLR)	642.938***	638.608***	895.560***	519.209***
df	464	454	588	379
CFI	0.944	0.943	0.912	0.955
RMSEA	0.050	0.051	0.057	0.048
SRMR	0.053	0.050	0.074	0.049
GFI	0.981	0.981	0.975	0.984
AGFI	0.975	0.975	0.968	0.980
N	208	208	182	208

† p < .1; * p < .05; ** p < .01; *** p < .001
* SDA: self-disclosure accuracy, SDC: self-disclosure concerns, PrC: privacy concerns, T: trust, ABD: awareness of big data, ABDI: awareness of big data implications, PeC: perceived control, PV: perceived vulnerability, SE: self-efficacy

Model 3 replicated Model 1 while incorporating four demographic variables (gender, age, education, and ethnicity) as control variables of the outcome constructs (SDC and SDA). The approximate fit indices indicated that this model demonstrated good fit (GFI = 0.975, AGFI = 0.968, CFI = 0.912, RMSEA = 0.057, SRMR = 0.074). In terms of the significance of the path coefficients, nothing differed compared to Model 1's findings except that self-efficacy was not any more significant even at the 0.1 significance level. In addition, the results from this model indicate that none of the demographic variables was significantly associated with the outcome constructs. However, only age was positively and significantly associated with self-disclosure accuracy (age → DV estimate = 041; se = 0.021). This finding suggests that older users tend to provide accurate data compared to younger users. We conducted a chi-square difference test to compare this model with Model 1. Model 3's sample size (N = 182) differed from Model 1's (N = 208); therefore, we needed to adjust Model 1's sample size to correctly compute this test. After correcting for sample size so that both models have equal sample sizes (N = 182), the results indicated that Model 1 demonstrated significantly better fit than Model 3 ($\chi_{(ML)}$ diff = 255.88; $p < 0.000$).

Model 4 replicated Model 1 while treating the trust construct as a predictor of PrC, SDC, and SDA. The approximate fit indices indicated that this model demonstrated good fit (GFI = 0.984, AGFI = 0.980, CFI = 0.955, RMSEA = 0.048, SRMR = 0.049). Similarly, the results did not differ compared to Model 1 except for self-efficacy. The results show that trust was not significantly associated with PrC, SDC, or SDA at the 0.05 significance level (T → PrC estimate = -0.034; se = 0.072; T → SDC estimate = -0.157; se = 0.088; T → SDA estimate = 0.049; se = 0.121). The directions of these estimates support the notion that trust in social networking websites decreases privacy concerns and increases the accuracy of data that users provide. However, none of these estimates were significant except the path from trust to self-disclosure concerns, which was significant only at the 0.1 level. We conducted a chi-square difference test to test whether Model 1 and Model 4 exhibited a difference in terms of model fit. The result from this test indicated a significant difference between the two models ($\chi_{(ML)}$ diff = 123.42; $p = 0.004$). This finding suggests that Model 4 had a better fit compared to Model 1. We also tested other alternative models treating trust as an outcome of privacy concerns and as a mediator between privacy concerns and the final outcomes (Malhotra et al., 2004). The results (not presented) indicated no significant improvements in terms of model fit, no significant path coefficients that pertained to trust, and no differences compared to the results found in Model 1. Table 4 summarizes the findings in terms of support for the proposed hypotheses.

Table 4. Summary of Results

Hypothesis	Model 1	Model 2	Model 3	Model 4
Hypothesis 1: Privacy concerns are positively associated with self-disclosure concerns.	Supported	Supported	Supported	Supported
Hypothesis 2: Privacy concerns are negatively associated with self-disclosure accuracy.	Supported	Supported	Supported	Supported
Hypothesis 3: Trust moderates the negative relationship between privacy concerns and self-disclosure accuracy such that the relationship is weaker (stronger) when trust is high (low).	Not supported	Not supported	Not supported	-
Hypothesis 4: Awareness of big data is negatively associated with privacy concerns.	Supported	Supported	Supported	Supported
Hypothesis 5: Awareness of big data implications is positively associated with privacy concerns.	Supported	Supported	Supported	Supported
Hypothesis 6: Perceived control is negatively associated with privacy concerns.	Supported	Supported	Supported	Supported
Hypothesis 7: Perceived vulnerability is positively associated with privacy concerns.	Supported	Supported	Supported	Supported
Hypothesis 8: Self-efficacy is negatively associated with privacy concerns.	Supported †	Supported †	Not supported	Not supported
† Supported only at the 0.1 significance level.				

6 Discussion

In this study, we examine how familiarity with big data impacts privacy concerns and disclosure outcomes among users of social networking websites. We also extend privacy research in the area of social networking websites by accounting for important factors unexamined in previous research (see Table A3). By using structural equation modeling techniques, we found support for most of the proposed hypotheses and the mediating role of privacy concerns in an APCO model.

The findings contend that awareness of big data and big data implications (collection, storing, processing, and sharing) have an impact on users' privacy concerns. On the one hand, users who are aware of the meaning of big data and their benefits tend to have less concern. On the other hand, users express more concerns as they become aware of big data implications. In line with the literature, we also found that privacy concerns significantly influence users' intentions to provide inaccurate and incomplete information. In particular, concerned users are more willing to falsify their personal data or refuse to give them out. Further, we found that negative attitudes toward privacy associate positively with self-disclosure concerns, an attitude that reflects concerns about the amount of personal information revealed to social networking websites.

Consistent with PMT's and TPB's predictions, our results show that perceived control, perceived vulnerability, and self-efficacy are significantly associated with privacy concerns. The findings indicate that users who perceive that they have control over the data they submit to social networking websites tend to be less concerned. In addition, users who believe that they are able to cope with privacy threats tend to be less concerned. However, those who feel vulnerable to and suspicious about potential risks tend to be more concerned. To summarize, our study shows that several antecedents whose effect on disclosure outcomes manifests indirectly through privacy concerns can explain privacy concerns.

These findings provide important practical implications for big data practitioners and social media providers. Practitioners need to be aware that big datasets obtained from social media contain some portions of false and incomplete data and users' privacy concerns are one reason for this accuracy issue. We encourage social media providers to improve their information-handling procedures, assurance mechanisms, and privacy settings in order to lessen users' concerns and their subsequent undesirable consequences (Bansal et al., 2015; Lowry et al., 2012; Posey et al., 2010; Wu et al., 2012). This study shows that being aware of big data and its value to individuals and businesses is associated with lower levels of privacy concerns. Thus, we advise social media providers to inform users about the potential benefits of big data and its implications as another way to mitigate privacy concerns and enhance the accuracy of data that users generate.

Several social networking websites do not offer sufficient level of control to their users (Wisniewski et al., 2016). Facebook is an exception and offers its users with a broad set of privacy settings (Cavusoglu et al., 2016). Users can employ these settings in order to manage their privacy and to have some control over what they share on the network. Nevertheless, research has shown that many users do not use privacy settings (Wisniewski et al., 2016) mainly because they do not know about these settings or they find them difficult to use. Self-efficacy and perceived control are important factors that can mitigate users' privacy concerns. As a result, social networking websites should offer a broad set of privacy settings that are easy to use so that users perceive higher level of self-efficacy and control because research has already shown the importance of privacy settings' ease of use in terms of determining the actual use of these settings (Keith et al., 2014).

While most of our findings concur with previous research conducted in other contexts such as the Internet and online retail websites (Dinev & Hart, 2004; Sheehan & Hoy, 1999; Son & Kim, 2008; Yao et al., 2007), some results contradict previous research. For instance, Mohamed and Ahmed (2012) found that self-efficacy is positively associated with privacy concerns. As we discuss in Section 3.6, we expected a negative relationship between self-efficacy and privacy concerns because it is more logical to think that users who have capabilities to cope with privacy threats would be less concerned. Although we adapted Mohamed and Ahmed's (2012) self-efficacy measurement, the findings still support our rationale and are in line with other research (Boss et al., 2015; Crossler & Bélanger, 2009, 2014; Keith et al., 2015). Several confounds can explain this contradiction (e.g., sample characteristics and privacy concerns measurements), but we call on future researchers to examine this relationship. With regard to demographic variables, our findings show that older users are more willing to disclose accurate personal information than younger users. Still, one should interpret this finding with caution because our sample did not include many users above 36 years of age. We found no differences with regard to gender, level of

education, and ethnicity. Nevertheless, when we included demographic variables in our models, the model fit attenuated significantly. Hence, future researchers should be vigilant of incorporating demographic variables into a complex research model.

Unexpectedly, we found no support for the mitigating role of trust, which suggests that trust in social networking websites may not be sufficient in lessening the negative effect of privacy concerns on self-disclosure accuracy. We further examined different research models in order to provide theoretical insights about modeling trust and the mediating role of privacy concerns. The results show that one can best model trust as a predictor of privacy concerns and its outcomes rather than a moderator or an outcome of privacy concerns. However, this analysis used only one sample, and, therefore, we encourage future researchers to investigate this relationship while providing a priori theoretical justification for such treatment. We also suggest future research to consider both trust and distrust as they reflect two different constructs that could influence privacy concerns in different ways (Moody, Galleta, & Lowry, 2014; Moody, Lowry, & Galletta, forthcoming). Finally, the results provide support for a fully mediated APCO model. Privacy concerns fully mediated the effects of the antecedents on self-disclosure accuracy and self-disclosure concerns (Smith et al., 2011).

7 Limitations and Future Research

This study has several limitations that one must consider when interpreting the results. We adopted a cross-sectional design, so one should interpret the results as associations rather than causations. We need future research that uses experimental designs. For example, it is possible to manipulate awareness of big data and its implications in order to make causal inferences about their exogenous effect on privacy concerns and disclosure outcomes. Recent research by Keith and colleagues is insightful in terms of using mobile apps to design longitudinal privacy experiments (Keith et al., 2014, 2015, 2016; Steinbart, Keith, & Babb, 2016). It would also be interesting to investigate the interaction between familiarity with big data and assurance mechanisms in social networking and retailing websites. We need to understand which of these constructs becomes more salient when users make disclosure and purchasing decisions. Such investigation will help organizations in redesigning their assurance mechanisms (e.g., privacy statements and privacy seals) to highlight the most effective message to render positive users' behaviors (Lowry et al., 2011).

Further, we used a purposive sampling method based on one U.S. academic institution. Accordingly, our findings have limited external validity. As such, we need future research that uses other sampling methods to extend the generalizability of our findings. For instance, future research can use Amazon Mechanical Turk (MTurk) to further test our framework. As Lowry, D'Arcy, Hammer, and Moody (2016) note, MTurk is "a compelling new source of data, MTurk leverages the crowdsourcing model that enables new means of accessing and filtering that were previously impossible" (p. 234). IS researchers are increasingly using MTurk because it has several advantages over traditional data-collection methods (Lowry et al., 2016; Steinbart et al., 2017; Trinkle et al., 2015).

Even though we accounted for several factors, it could be possible that other unobserved variables confounded the results. We integrated PMT and TPB in order to test an APCO model. However, we did not account for several factors that PMT and TPB suggest. For instance, PMT suggests that perceived severity, rewards, response efficacy, and response cost are important predictors of fear appeals and protective behaviors. TPB also suggests that subjective norms toward the issue of private information represent an important predictor of attitudes toward privacy. We did not consider these factors in the research model, and, as such, we clearly need future research to test these theories in their full extent. IS security research has already provided strong support for the predictive power of PMT (Boss et al., 2015). However, we lack a thorough test of PMT in the privacy domain. Thus, IS privacy researchers can contribute theoretically to PMT by testing its nomology.

Another interesting avenue would include more thoroughly investigating how trust plays a role in the context of social networking websites. As we discuss above, trust is a complex latent construct. We treated trust as a unidimensional construct based on published privacy research. However, Moody et al. (2014, forthcoming) found that trust and distrust coexist in online contexts and result in ambivalence when both have attitudinal values. They also found that distrust has a stronger effect on intentions than does trust (Moody et al., 2014) and that one needs to use advanced statistical techniques (e.g., polynomial regression analysis) to understand their effect on users' intentions (Moody et al., forthcoming). This area of research is still burgeoning, but we strongly recommend future privacy research to account for both

trust and distrust in their research frameworks because such bidimensional perspective could reveal novel insights on the role of trust in privacy-disclosure contexts.

The measurements we used are not tied to a specific social networking website. As such, users could exhibit different attitudes and behaviors when using different websites (e.g., Facebook, Twitter, LinkedIn, and Instagram). Users' level of trust and disclosure behaviors may actually depend on their expectations from each social network (Burgoon et al., 2016). Future research could examine privacy concerns and disclosure outcomes across different social networking websites. Last, we measured our main dependent variable (i.e., self-disclosure accuracy) based on intentions to falsify and refusal to give out personal information. We did not measure actual behaviors. Although recent research contends that intentions strongly predict actual behaviors (Boss et al., 2015; Keith et al., 2013), we need future research to consider actual disclosure behaviors in addition to intentions.

8 Concluding Remarks

One of the most appealing benefits of big data is monetized analytics (Schmarzo, 2013). Social networking websites adopt monetized analytics to generate additional revenues by collecting and analyzing a massive amount of data that they eventually commercialize to interested parties (Craig & Ludloff, 2011; Hurwitz et al., 2013). However, unreliable big datasets are destructive to both social networking websites and data brokers because both run the risk of using poor quality data. Incorrect information leads practitioners to make poor decisions, and, with an insufficient amount of information, they cannot make sound decisions (Sheehan & Hoy, 1999). The results from this study should prompt social networking websites to enhance their privacy practices, which they can achieve by enhancing the user-friendliness of privacy controls and emphasizing the potential benefits of the use of personal data through explicit statements in privacy policies. By doing so, they can reduce users' privacy concerns and eventually improve the quality of data that users generate.

We are in an era of big data where online companies collect, store, process, and share personal data for various purposes (Boyd & Crawford, 2012; Craig & Ludloff, 2011; Shim et al., 2015; Watson, 2014). Yet, we lack privacy research that explores factors that pertain to individuals' awareness of this digital era. Awareness of big data will likely influence individuals' attitudes toward privacy and their disclosure behaviors. IS privacy theories need to consider these evolutionary changes in order to provide insights on how the era of big data has changed individuals' understanding of information privacy and disclosure behaviors.

Acknowledgments

This paper is based on the master's thesis of the first author, which the third author advised. We thank the associate editor for the insightful comments on earlier versions of this manuscript. We also thank Emma Vigneault and Richard Vigneault who assisted in proofreading the manuscript.

References

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2), i-xxxii.

Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the facebook. In G. Danezis & P. Golle (Ed.), *Proceedings of the 6th Workshop on Privacy enhancing technology* (pp. 36-58). Cambridge, UK: Robinson College.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211.

Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. *Psychology & Health*, *26*(9), 1113-1127.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, New Jersey: Prentice-Hall.

Alashoor, T, Lambert, L., & Farivar, S. (2016). A review of measures of disclosure outcomes in the IS privacy literature. In *Proceedings of the 22nd Americas Conference on Information Systems.*

Alashoor, T., & Baskerville, R. (2015). The privacy paradox: The role of cognitive absorption in the social networking activity. In *Proceedings of the 36th International Conference on Information Systems.*

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly, 26*(3), 243-268.

Bandura, A. (1988). Self-efficacy conception of anxiety. *Anxiety Research*, *1*(2), 77-98.

Bansal, G., Zahedi, F. M., & Gefen, D. (2015). The role of privacy assurance mechanisms in building trust and the moderating role of privacy concern. *European Journal of Information Systems*, *24*(6), 624-644.

Bansal, G., Zahedi, F. M., & Gefen, D. (2016). Do context and personality matter? Trust and privacy concerns in disclosing private information online. *Information & Management*, *53*, 1-21.

Bansal, G., Zahedi, F., & Gefen, D. (2008). The moderating influence of privacy concern on the efficacy of privacy assurance mechanisms for building trust: A multiple-context investigation. In *Proceedings of 29th International Conference on Information Systems*.

Bapna, R., Gupta, A., Rice, S., & Sundararajan, A. (forthcoming). Trust and the strength of ties in online social networks: An exploratory field experiment. *MIS Quarterly*.

Baruh, L., & Cemalcılar, Z. (2014). It is more than personal: Development and validation of a multidimensional privacy orientation scale. *Personality and Individual Differences, 70*, 165-170.

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly, 35*(4), 1017-1041.

Belanger, F., Hiller, J. S., & Smith, W. J. (2002). Trustworthiness in electronic commerce: The role of privacy, security, and site attributes. *The Journal of Strategic Information Systems*, *11*(3), 245-270.

Bertolucci, J. (2013). Privacy concerns: Big data's biggest barrier? *Informationweek*. Retrieved from http://businessintelligence.com/bi-insights/privacy-concerns-big-datas-biggest-barrier/

Beyer, M. A., & Laney, D. (2012). The importance of “big data”: A definition. *Gartner.* Retrieved from https://www.gartner.com/doc/2057415/importance-big-data-definition

Boss, S. R., Galletta, D. F., Lowry, P. B., Moody, G. D., & Polak, P. (2015). What do users have to fear? Using fear appeals to engender threats and fear that motivate protective security behaviors. *MIS Quarterly*, *39*(4), 837-864.

Boyd, D. (2004). Friendster and publicly articulated social networking. In *Proceedings of the Conference on Human Factors and Computing Systems* (pp. 1279-1282).

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662-679.

Breznitz, D., Murphree, M., & Goodman, S. (2011). Ubiquitous data collection: Rethinking privacy debates. *Computer, 44*(6), 100-102.

Burgoon, J. K., Bonito, J. A., Lowry, P. B., Humpherys, S. L., Moody, G. D., Gaskin, J. E., & Giboney, J. S. (2016). Application of expectancy violations theory to communication with and judgments about embodied agents during a decision-making task. *International Journal of Human-Computer Studies*, *91*, 24-36.

Cao, J., Basoglu, K. A., Sheng, H., & Lowry, P. B. (2015). A Systematic review of social networks research in information systems: Building a foundation for exciting future research. *Communications of the Association for Information Systems*, *36,* 727-758.

Cavusoglu, H., Phan, T. Q., Cavusoglu, H., & Airoldi, E. M. (2016). Assessing the impact of granular privacy controls on content sharing and disclosure on Facebook. *Information Systems Research*, *27*(4), 848-879.

Cazier, J. A., Jensen, A. S., & Dave, D. S. (2008). The impact of consumer perceptions of information privacy and security risks on the adoption of residual RFID technologies. *Communications of the Association for Information Systems*, *23*, 235-256.

Chan, Y. E., Culnan, M. J., Greenaway, K., Laden, G., Levin, T., & Smith, H. J. (2005). Information privacy: Management, marketplace, and legal challenges. *Communications of the Association for Information Systems*, *16*, 270-298.

Chen, K., & Rea, A. I. (2004). Protecting personal information online: A survey of user privacy concerns and control techniques. *The Journal of Computer Information Systems, 44*(4), 85-92.

Chen, R., & Sharma, S. K. (2013). Understanding member use of social networking sites: A value analysis. *Communications of the Association for Information Systems*, *33*(7), 97-114.

Choi, B. C., & Land, L. (2016). The effects of general privacy concerns and transactional privacy concerns on Facebook apps usage. *Information & Management*, *53*, 868-877

Choi, B. C., Jiang, Z., Xiao, B., & Kim, S. S. (2015). Embarrassing exposures in online social networks: An integrated perspective of privacy invasion and relationship bonding. *Information Systems Research*, *26*(4), 675-694.

Clemons, E. K., Wilson, J., & Jin, F. (2014). Investigations into consumers preferences concerning privacy: An initial step towards the development of modern and consistent privacy protections around the globe. In *Proceedings of the 47th Hawaii International Conference on System Sciences (*pp. 4083-4092).

Craig, T., & Ludloff, M. E. (2011). *Privacy and big data.* Sebastopol, CA: O'Reilly Media.

Crossler, R. E., & Bélanger, F. (2009). The effects of security education training and awareness programs and individual characteristics on end user security tool usage. *Journal of Information System Security*, *5*(3), 3-22.

Crossler, R., & Bélanger, F. (2014). An extended perspective on individual security behaviors: Protection motivation theory and a unified security practices (USP) instrument. *The Data Base for Advances in Information Systems*, *45*(4), 51-71.

Dinev, T., & Hart, P. (2005). Internet privacy concerns and social awareness as determinants of intention to transact. *International Journal of Electronic Commerce*, *10*(2), 7-29.

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research, 17*(1), 61-80.

Dinev, T., Hart, P. (2004). Internet privacy concerns and their antecedents—measurement validity and a regression model. *Behaviour & Information Technology, 23*(6), 413-422.

Drake, J. R., Hall, D., Becton, J. B., & Posey, C. (2016). Job applicants' information privacy protection responses: Using social media for candidate screening. *AIS Transactions on Human-Computer Interaction*, *8*(4), 160-184.

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). Frequency of social media use. *The Pew Research Center*. Retrieved from http://www.pewinternet.org/2015/01/09/frequency-of-social-media-use-2/

Fernback, J., & Papacharissi, Z. (2007). Online privacy as legal safeguard: The relationship among consumer, online portal, and privacy policies. *New Media & Society, 9*(5), 715-734.

Floyd, D. L., Prentice-Dunn, S., & Rogers, R. W. (2000). A meta-analysis of research on protection motivation theory. *Journal of Applied Social Psychology*, *30*(2), 407-429.

Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior, 25*(1), 153-160.

Gantz, J. F., Chute, C., Manfrediz, A., Minton, S., Reinsel, D., Schlichting, W., & Toncheva, A. (2008). *An update forecast of worldwide information growth through 2011*. IDC.

Gao, G. (2015). What Americans think about NSA surveillance, national security and privacy. *Pew Research Center*. Retrieved from http://www.pewresearch.org/fact-tank/2015/05/29/what-americans-think-about-nsa-surveillance-national-security-and-privacy/

Gefen, D., Rao, V. S., & Tractinsky, N. (2003). The conceptualization of trust, risk and their electronic commerce: The need for clarifications. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*.

GfK (2014). GfK survey on data privacy and trust. *Growth From Knowledge*. Retrieved from http://docplayer.net/9299896-Gfk-survey-on-data-privacy-and-trust.html

Gundecha, P., & Liu, H. (2012). Mining social media: A brief introduction. *Tutorials in Operations Research*, *1*(4), 1-17.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International journal of Information Management, 33*(3), 464-472.

Hong, W., & Thong, J. Y. L. (2013). Internet privacy concerns: An integrated conceptualization and four empirical studies. *MIS Quarterly, 37*(1), 275-298.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Hurwitz, J., Nugent, A., Halper, F., & Kaufman, M. (2013). *Big data for dummies*. Hoboken, NJ: John Wiley and Sons.

IBM. (2012). *Analytics: The real-world use of big data.* Retrieved from http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html

Jiang, Z., Heng, C. S., & Choi, B. C. (2013). Research note—privacy concerns and privacy-protective behavior in synchronous online social interactions. *Information Systems Research, 24*(3), 579-595.

Junglas, I. A., Johnson, N. A., & Spitzmüller, C. (2008). Personality traits and concern for privacy: An empirical study in the context of location-based services. *European Journal of Information Systems*, *17*(4), 387-402.

Kart, L., Heudecker, N., & Buytendijk, F. (2013). Survey analysis: Big data adoption in 2013 shows substance behind the hype. *Gartner.* Retrieved from https://www.gartner.com/doc/2589121/survey-analysis-big-data-adoption

Keith, M. J., Babb, J. S., & Lowry, P. B. (2014). A longitudinal study of information privacy on mobile devices. In *Proceedings of the 47th Hawaii International Conference on System* Sciences (pp. 3149-3158).

Keith, M. J., Babb, J. S., Lowry, P. B., Furner, C. P., & Abdullat, A. (2015). The role of mobile-computing self-efficacy in consumer information disclosure. *Information Systems Journal*, *25*(6), 637-667.

Keith, M. J., Babb, J., Furner, C., Abdullat, A., & Lowry, P. B. (2016). Limited information and quick decisions: Consumer privacy calculus for mobile applications. *AIS Transactions on Human-Computer Interaction, 8*(3), 88-130.

Keith, M. J., Maynes, C., Lowry, P. B., & Babb, J. (2014). Privacy fatigue: The effect of privacy control complexity on consumer electronic information disclosure. In *Proceedings of International Conference on Information Systems.*

Keith, M. J., Thompson, S. C., Hale, J., Lowry, P. B., & Greer, C. (2013). Information disclosure on mobile devices: Re-examining privacy calculus with actual user behavior. *International Journal of Human - Computer Studies, 71*(12), 1163-1173.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford.

Krasnova, H., Gunther, O., Spiekermann, S., & Koroleva, K. (2009). Privacy concerns and identity in online social networks. *Identity in the Information Society, 2*(1), 39-63.

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online social networks: Why we disclose. *Journal of Information Technology*, *25*(2), 109-125.

Krasnova, H., Veltri, N. F., & El Garah, W. (2014). Effectiveness of justice-based measures in managing trust and privacy concerns on social networking sites: An intercultural perspective. *Communications of the Association for Information Systems*, *35*, 83-108.

Krasnova, H., Veltri, N. F., and Günther, O. (2012). Self-disclosure and privacy calculus on social networking sites: The role of culture: Intercultural dynamics of privacy calculus. *Business and Information Systems Engineering,* 4(3), 127-135.

Kuo, K. M., Ma, C. C., & Alexander, J. W. (2014). How do patients respond to violation of their information privacy? *Health Information Management Journal, 43*(2), 23-33.

Li, H., Gupta, A., Zhang, J., & Sarathy, R. (2014). Examining the decision to use standalone personal health record systems as a trust-enabled fair social contract. *Decision Support Systems*, *57*, 376-386.

Li, T., & Slee, T. (2014). The effects of information privacy concerns on digitizing personal health records. *Journal of the Association for Information Science and Technology*, *65*(8), 1541-1554.

Li, Y. (2011). Empirical studies on online information privacy concerns: Literature review and an integrative framework. *Communications of the Association for Information Systems, 28*(1), 453-496.

Li, Y. (2012). Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems, 54*(1), 471-481.

Li, Y. (2014). The impact of disposition to privacy, website reputation and website familiarity on information privacy concerns. *Decision Support Systems*, *57*, 343-354.

Liu, C., Marchewka, J. T., Lu, J., & Yu, C. S. (2005). Beyond concern—a privacy-trust-behavioral intention model of electronic commerce. *Information & Management*, *42*(2), 289-304.

Lowry, P. B., Cao, J., & Everard, A. (2011). Privacy concerns versus desire for interpersonal awareness in driving the use of self-disclosure technologies: The case of instant messaging in two cultures. *Journal of Management Information Systems*, *27*(4), 163-200.

Lowry, P. B., D'Arcy, J., Hammer, B., & Moody, G. D. (2016). "Cargo cult" science in traditional organization and information systems survey research: A case for using nontraditional methods of data collection, including Mechanical Turk and online panels. *The Journal of Strategic Information Systems*, *25*(3), 232-240.

Lowry, P. B., Moody, G., Vance, A., Jensen, M., Jenkins, J., & Wells, T. (2012). Using an elaboration likelihood approach to better understand the persuasiveness of website privacy assurance cues for online consumers. *Journal of the American Society for Information Science and Technology*, *63*(4), 755-776.

Lowry, P. B., Roberts, T., & Caine, B. (2005). Familiarity effects on trust with mobile computing device websites. In *Proceedings of the 11th International Conference on HCI.*

Lowry, P. B., Vance, A., Moody, G., Beckman, B., & Read, A. (2008). Explaining and predicting the impact of branding alliances and web site quality on initial consumer trust of e-commerce web sites. *Journal of Management Information Systems*, *24*(4), 199-224.

Lowry, P. B., Wilson, D. W., & Haig, W. L. (2014). A picture is worth a thousand words: Source credibility theory applied to logo and website design for heightened credibility and consumer trust. *International Journal of Human-Computer Interaction*, *30*(1), 63-93.

Lwin, M. O., & Williams, J. D. (2003). A model integrating the multidimensional developmental theory of privacy and theory of planned behavior to examine fabrication of information online. *Marketing Letters, 14*(4), 257-272.

Maddux, J. E., & Rogers, R. W. (1983). Protection motivation theory and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of Experimental Social Psychology, 19*(5), 469-479.

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.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey*. Retrieved from http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Mishra, S. (2015). Adoption of m-commerce in India: Applying theory of planned behaviour model. *The Journal of Internet Banking and Commerce*, *19*(1), 1-17.

Mohamed, A. A. (2010). Online privacy concerns among social networks' users. *Cross-Cultural Communication, 6*(4), 74-89.

Mohamed, N., & Ahmad, I. H. (2012). Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia. *Computers in Human Behavior, 28*(6), 2366-2375.

Moody, G. D., Galletta, D. F., & Lowry, P. B. (2014). When trust and distrust collide online: The engenderment and role of consumer ambivalence in online consumer behavior. *Electronic Commerce Research and Applications*, *13*(4), 266-282.

Moody, G. D., Lowry, P. B., & Galletta, D. F. (Forthcoming). It's complicated: Explaining the relationship between trust, distrust, and ambivalence in online transaction relationships using polynomial regression analysis and response surface analysis. *European Journal of Information Systems*.

Pavlou, P. A., & Fygenson, M. (2006). Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior. *MIS Quarterly, 30*(1), 115-143.

Ponemon (2014a). Is your company ready for a big data breach? The second annual study on data breach preparedness. *Ponemon Institute*. Retrieved from http://www.ponemon.org/blog/is-your-company-ready-for-a-big-data-breach-the-second-annual-study-on-data-breach-preparedness

Ponemon (2014b). 2014 Cost of data breach: Global. *Ponemon Institute*. Retrieved from http://www.ponemon.org/library/2014-cost-of-data-breach-global?s=%243.5+million

Posey, C., Bennett, B., Roberts, T., & Lowry, P. B. (2011). When computer monitoring backfires: Invasion of privacy and organizational injustice as precursors to computer abuse. *Journal of Information System Security, 7*(1), 24-47

Posey, C., Lowry, P. B., Roberts, T. L., & Ellis, T. S. (2010). Proposing the online community self-disclosure model: The case of working professionals in France and the UK who use online communities. *European Journal of Information Systems*, *19*(2), 181-195.

Rainie, L., Kiesler, S., Kang, R., & Madden, M. (2013). Anonymity, privacy, and security online. *The Pew Research Center*. Retrieved from http://www.pewinternet.org/2013/09/05/anonymity-privacy-and-security-online/

Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change. *Journal of Psychology, 91*(1), 93-114.

Russom, P. (2011). Big data analytics. *TDWI.* Retrieved from http://tdwi.org/portals/big-data-analytics.aspx

Schmarzo, B. (2013). *Big data: Understanding how data powers big business*. Indianapolis, IN: John Wiley & Sons.

Sharma, S., & Crossler, R. E. (2014). Disclosing too much? Situational factors affecting information disclosure in social commerce environment. *Electronic Commerce Research and Applications*, *13*(5), 305-319.

Shey, H. (2013). Understand the state of data security and privacy: 2013 to 2014. *Forrester Research.* Retrieved from https://www.forrester.com/report/Understand+The+State+Of+Data+Security+And+Privacy+2013+To+2014/-/E-RES82021

Sheehan, K. B., & Hoy, M. G. (1999). Flaming, complaining, abstaining: How online users respond to privacy concerns. *Journal of Advertising, 28*(3), 37-52.

Sheng, H., Nah, F. F. H., & Siau, K. (2008). An experimental study on ubiquitous commerce adoption: Impact of personalization and privacy concerns. *Journal of the Association for Information Systems*, *9*(6), 344-376.

Shim, J. P., Dekleva, S., French, A. M., & Guo, C. (2013). Social networking and social media in the United States, South Korea, and China. *Communications of the Association for Information Systems*, *33*, 485-496.

Shim, J. P., French, A. M., Guo, C., & Jablonski, J. (2015). Big data and analytics: Issues, solutions, and ROI. *Communications of the Association for Information Systems*, *37*, 797-810.

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly, 35*(4), 989-1015.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information privacy: Measuring individuals' concerns about organizational practices. *MIS Quarterly, 20*(2), 167-196.

Son, J., & Kim, S. S. (2008). Internet users' information privacy-protective responses: Taxonomy and a nomological model. *MIS Quarterly, 32*(3), 503-529.

Steinbart, P. J., Keith, M. J., & Babb, J. (2016). Examining the continuance of secure behavior: A longitudinal field study of mobile device authentication. *Information Systems Research*, *27*(2), 219-239.

Steinbart, P., Keith, M., & Babb, J. (2017). Measuring privacy concern and the right to be forgotten. In *Proceedings of the 50th Hawaii International Conference on System Sciences* (pp. 4967-4976).

Stewart, K. A., & Segars, A. H. (2002). An empirical examination of the concern for information privacy instrument. *Information Systems Research, 13*(1), 36-49.

Taddei, S., & Contena, B. (2013). Privacy, trust and control: Which relationships with online self-disclosure? *Computers in Human Behavior, 29*(3), 821-826.

Tan, W., Blake, M. B., Saleh, I., & Dustdar, S. (2013). Social-network-sourced big data analytics. *IEEE Internet Computing, 17*(5), 62-69.

Tow, W. N. F. H., Dell, P., & Venable, J. (2010). Understanding information disclosure behaviour in Australian Facebook users. *Journal of Information Technology*, *25*(2), 126-136.

Trinkle, B. S., Crossler, R. E., & Bélanger, F. (2015). Voluntary disclosures via social media and the role of comments. *Journal of Information Systems*, *29*(3), 101-121.

Turow, J., & Hennessy, M. (2007). Internet privacy and institutional trust: Insights from a national survey. *New Media & Society, 9*(2), 300-318.

Varadarajan, S., & Soundarapandian. (2013). Maximizing insight from unstructured data. *Business Intelligence Journal, 18*(3), 17-25.

Vera-Baquero, A., Colomo-Palacios, R., & Molloy, O. (2013). Business process analytics using a big data approach. *IT Professional, 15*(6), 29-35.

Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, *34*, 1247-1268.

Wisniewski, P., Islam, A. K. M., Lipford, H. R., & Wilson, D. C. (2016). Framing and measuring multi-dimensional interpersonal privacy preferences of social networking site users. *Communications of the Association for Information Systems*, *38*, 235-258.

Wu, K. W., Huang, S. Y., Yen, D. C., & Popova, I. (2012). The effect of online privacy policy on consumer privacy concern and trust. *Computers in Human Behavior*, *28*(3), 889-897.

Xie, E., Teo, H., & Wan, W. (2006). Volunteering personal information on the Internet: Effects of reputation, privacy notices, and rewards on online consumer behavior. *Marketing Letters, 17*(1), 61-74.

Xu, F., Michael, K., & Chen, X. (2013). Factors affecting privacy disclosure on social network sites: an integrated model. *Electronic Commerce Research*, *13*(2), 151-168.

Yao, M. Z., Rice, R. E., & Wallis, K. (2007). Predicting user concerns about online privacy. *Journal of the American Society for Information Science and Technology, 58*(5), 710-722.

Youn, S. (2009). Determinants of online privacy concern and its influence on privacy protection behaviors among young adolescents. *The Journal of Consumer Affairs, 43*(3), 389-418.

Appendix A: Review of Privacy Research

Table A1. Privacy Research Adopting/Adapting Protection Motivation Theory (PMT)

Citation	Context	Independent variables	Mediators	Dependent variables	Theories used	Summary of significant results
Boss et al. (2015) Study 2	Anti-malware software	Perceived severity, perceived vulnerability, rewards, response efficacy, self-efficacy, and response cost.	Fear appeals and anti-malware software use intention	Anti-malware software use	PMT	High fear appeal condition: All PMT predictions hold. Fear appeals partially mediate the relationships between threat appraisals constructs and anti-malware software use intention. Threat appraisals (severity and vulnerability) are positively associated with intention. Threat appraisal (rewards) is negatively associated with intention. Coping appraisals are positively associated with intention, except that response cost is negatively associated with intention. Intention is positively associated with actual use. Low fear appeal condition: the results contradict PMT in several aspects.
Crossler & Bélanger (2014)	Home computers and networks	perceived severity, perceived vulnerability, response efficacy, self-efficacy, and response cost		Unified security practices	PMT	Perceived severity, response efficacy, and self-efficacy are positively associated with security practices. Perceived vulnerability is negatively associated with security practices.
Dinev & Hart (2004)	Internet	Perceived vulnerability and perceived control		Perceived privacy concerns	PMT†	Perceived vulnerability is positively associated with privacy concerns.
Junglas et al. (2008)	Location-based services	Big five personality traits		Concern for privacy	PMT	Conscientiousness and openness to experience are positively associated with concern for privacy. Agreeableness is negatively associated with concern for privacy.
Kuo et al. (2014)	Electronic medical records (ERDs)	Privacy concerns dimensions: collection, unauthorized access, secondary use, and errors		Information privacy-protective responses	PMT	Collection, secondary use, and errors are positively associated with protective responses (e.g., intentions to refuse to provide personal health information and misrepresentation of personal health information).
Yao et al. (2007)	Internet	Need for privacy, self-efficacy	Beliefs in privacy rights, Internet use fluency, and Internet use diversity	Concern about privacy	Social cognitive theory	Need for privacy is positively associated with concern about privacy directly and indirectly through beliefs in privacy rights. Self-efficacy is only indirectly associated with concern about privacy through the three mediators.
Youn (2009)*	Internet	Perceived vulnerability, perceived benefits, and self-efficacy	Privacy concerns	Privacy-protective behaviors (fabricate, seek advice, and refuse to use website)	PMT	Perceived vulnerability is positively and perceived benefit is negatively associated with privacy concerns. Privacy concern is positively associated with protective behaviors.

† The authors do not explicitly mention PMT although the independent variable (perceived vulnerability) represents PMT threat appraisals.
* Non-IS study

Table A2. Privacy Research Adopting/Adapting Theory of Planned Behavior (TPB)

Citation	Context	Independent variables	Mediators	Dependent variables	Theories used	Summary of significant results
Cazier et al. (2008)	RFID technologies	Perceived ease of use, perceived privacy risk likelihood, and perceived privacy risk harm	Perceived usefulness	Intention to use	TRA/TPB and technology acceptance model (TAM)	Both privacy-related constructs are negatively associated with intention to use.
Chen & Sharma (2013)	Social networking websites	Critical mass, social conformity, information quality, social presence, value of privacy, and Internet risk	Network management, social image, learning, enjoyment, risk, and attitude	Site use	TRA/TPB and theory of consumption value	Value of privacy and Internet risk are positively associated with risk (a conditional value). Risk is negatively associated with attitude toward a social networking website. Attitude is positively associated with website use.
Dinev & Hart (2005)	Internet	Internet literacy and social awareness	Privacy concerns	Intention to transact	TRA/TPB	Internet literacy is negatively associated with privacy concerns and positively associated with intention to transact. Social awareness is positively associated with privacy concerns. Privacy concern is negatively associated with intention to transact.
Dinev & Hart (2006)	Internet	Perceived Internet privacy risk (PR) and personal Internet interest (PI)	Internet privacy concerns (PC) and Internet trust (T)	Willingness to provide personal information to transact (PPIT)	Privacy calculus and TRA/TPB	PR is positively associated with PC and negatively associated with trust and PPIT. PC is negatively associated with PPIT. Trust is positively associated with PPIT. PI is positively associated with PPIT.
Keith et al. (2013)	Mobile devices	Privacy risk awareness, perceived benefits, and privacy concerns	Perceived privacy risks and intent to disclose	Actual disclosure and privacy settings	Privacy calculus and TRA/TPB	Privacy risk awareness is positively associated with perceived privacy risk. Privacy concern is positively associated with perceived privacy risk and negatively associated with intent to disclose. Perceived privacy risk is negatively associated with actual disclosure and privacy settings. Perceived benefit is positively associated with intent to disclose. Intent to disclose is positively associated with actual disclosure. Intent to disclose fully mediates the relationship between privacy concern and actual disclosure.
Keith et al. (2014)	Mobile game with social networking components	Benefits (manipulated app rewards), trust, and privacy concerns	Perceived benefits and perceived risk	Profile disclosure and profile accuracy	Privacy calculus, TRA/TPB, and prospect theory	Benefits are associated positively with perceived benefits, perceived risk, profile disclosure, and profile accuracy. The association between benefits and accuracy is stronger when trust is high (authors modeled trust as a moderator in this case). Trust is associated positively with profile disclosure and negatively with perceived risk. However, trust is not associated significantly with profile accuracy. Privacy concerns are associated positively with perceived risk and negatively with profile disclosure and profile accuracy. Perceived risk is associated negatively with profile accuracy.
Keith et al. (2015, Study 2)	Social networks	Mobile-computing self-efficacy (MCSE), privacy concern, and privacy settings	Perceived risk and perceived benefits	Actual disclosure	Privacy calculus, TRA/TPB, and social cognitive theory (SCT)	MCSE is negatively associated with perceived risk and positively associated perceived benefit. Privacy concern is positively associated with perceived risk. Perceived risk is negatively associated with disclosure and perceived benefit is positively associated with disclosure. Privacy settings use is negatively associated with disclosure.

Table A2. Privacy Research Adopting/Adapting Theory of Planned Behavior (TPB)

Li (2014)	Online retailers	Privacy experience, website reputation, website familiarity, and perceived benefits	Disposition to privacy and website-specific privacy concerns	Behavioral intention	TRA/TPB, Westin's and Altman's theories of privacy	Privacy experience is positively associated with disposition to privacy. Disposition to privacy is positively associated with website-specific privacy concerns. Website reputation and familiarity are negatively associated with website-specific privacy concerns. Website-specific privacy concern is negatively associated with behavioral intention. Perceived benefit is positively associated with behavioral intention.
Li et al. (2012)	Online retailers	Joy, fear, perceived relevance of information, awareness of privacy statement, and general privacy concern	Privacy protection belief and privacy risk belief	Behavioral intention	TRA/TPB, privacy calculus, and stimulus-organism-response model	Joy is positively associated with privacy protection belief and negatively associated with privacy risk belief. Fear is negatively associated with privacy protection belief and positively associated with privacy risk belief. Perceived relevance of information is positively associated with privacy protection belief and negatively associated with privacy risk belief. Awareness of privacy statement is positively associated with privacy protection belief. General privacy concern is positively associated with privacy risk belief and negatively associated with behavioral intention. Privacy protection belief is positively associated with behavioral intention. Privacy risk belief is negatively associated with behavioral intention.
Liu et al. (2005)	E-commerce	Privacy level	Trust	Behavioral intention	TRA/TPB	The higher the privacy protection provided by an e-commerce website, the higher the trust users have. Trust is positively associated with intention.
Lowry et al. (2011)	Instant Messaging (IM) technology	Hofstede's cultural dimensions (masculinity, uncertainty avoidance, power distance, and collectivism)	Information privacy concerns, desire for awareness, attitude toward IM technology, and behavioral intention to use	Use of IM	TRA/TPB and social exchange theory (SET)	Uncertainty avoidance and collectivism are positively associated with both information privacy concerns and desire for awareness, which, in turn, positively impacts attitudes toward IM technology. Attitudes toward IM technology are positively associated with behavioral intention to use IM, which, in turn, predicts use of IM positively.
Lowry et al. (2012)	E-commerce	Understanding privacy seals (PSs), sense of privacy seal assurance, perceived website transaction risk, negative media coverage, privacy victim, brand image, perceived website quality, presence of privacy assurance statements, and presence of privacy seals	Perceived privacy assurance	Behavioral intention toward website (e.g., giving information and making purchase)	Elaboration likelihood model (ELM) and TRA/TPB	Perceived website transaction risk and negative media coverage are negatively associated with perceived privacy assurance. Brand image, perceived website quality, and presence of privacy assurance statements are positively associated with perceived privacy assurance. Presence of privacy seals does not have a main effect on perceived privacy assurance but its interaction with PSs and sense of privacy seal assurance is positive and significant. Perceived privacy assurance is positively associated with behavioral intention toward website.
Malhotra et al. (2004)	Internet	Internet users' information privacy concerns (IUIPC)	Trusting beliefs and risk beliefs	Behavioral intentions	TRA/TPB	IUIPC is negatively associated with trusting beliefs and positively associated with risk beliefs. Trusting beliefs are positively associated with behavioral intention and negatively associated with risk beliefs. Risk beliefs are negatively associated with behavioral intention. More sensitive information request is associated with less behavioral intention.

Table A2. Privacy Research Adopting/Adapting Theory of Planned Behavior (TPB)

Sharma & Crossler (2014)	Social networks / commerce	Perceived fairness in exchange (surveillance, linkage, relevance), perceived enjoyment, perceived ownership, and privacy apathy	Perceived privacy risk and perceived usefulness	Intentions to disclose information	TRA/TPB, communication privacy management (CPM) theory, and fairness in exchange (FIE) theory	All perceived FIE constructs are associated with perceived privacy risk which in turn impacts intentions to disclose negatively. Privacy apathy is negatively associated with perceived privacy risk. Perceived enjoyment is positively associated with perceived usefulness. Privacy apathy, perceived enjoyment, and perceived usefulness are associated positively with intentions to disclose.
Son & Kim (2008)	Online companies	Information privacy concerns, perceived justice, and societal benefits		Information provision (refusal and misrepresentation), private action (removal and negative word-of-mouth), and public action (complaining directly to online companies and complaining indirectly to third party organizations)	TRA/TPB	Information privacy concern is positively associated with all dependent variables except misrepresentation. Perceived justice is negatively associated with both information provision constructs. Societal benefits are positively associated with both public action constructs.

Table A3. Privacy Research in the Context of Social Networking Websites

Study / related constructs	Familiarity with big data	Perceived control	Perceived vulnerability	Self-efficacy	Privacy concerns	Disclosure outcomes	Trust	Method
Baruh & Cemalcilar (2014)*					X	X		Survey
Choi & Land (2016)		X			X			Survey
Drake et al. (2016)					X	X		Survey
Fogel & Nehmad (2009)*		X			X	X	X	Survey
Jiang et al. (2013)					X	X		Survey
Keith et al. (2014)			X		X	X	X	Field experiment
Krasnova et al. (2009)					X	X		Survey & focus groups
Krasnova et al. (2010)		X			X	X	X	Survey
Krasnova et al. (2012)					X	X	X	Survey
Krasnova et al. (2014)					X		X	Survey
Mohamed (2010)*					X	X		Survey
Mohamed & Ahmad (2012)*			X	X	X			Survey
Posey et al. (2010)			X			X	X	Survey
Sharma & Crossler (2014)			X			X		Survey
Taddei & Contena (2013)*		X			X	X	X	Survey
Tow et al. (2010)					X	X	X	Survey & ethnography
Wisniewski et al. (2016)		X			X	X		Survey & interviews
Xu et al.(2013)		X			X	X		Survey
This study	X	X	X	X	X	X	X	Survey

* Non-IS study
Note: all studies used one or more social networking websites as the context. Jiang et al. (2013) used online chatting rooms as the context in. This table does not show the other constructs tested in the studies listed. It depicts only whether constructs related to this study were tested.

Table A4. Definitions of Constructs Tested in the this Study

Construct	Definition
Familiarity with big data	Awareness of the term big data, including its values to individuals and businesses, and the four major implications of big data (collection, storing, processing, and sharing).
Perceived control	The perceived ability to control personal data submitted to social networking websites.
Perceived vulnerability	The perceived potential risks associated with revealing personal information to social networking websites.
Self-efficacy	The confidence and competence to cope with potential privacy threats on social networking websites.
Privacy concerns	Concerns about opportunistic behavior related to the personal information submitted over social networking websites by the respondent in particular.
Trust	The extent to which users are confident that social networking websites will handle their personal information competently, reliably, and safely.
Self-disclosure concerns	Concerns about the extent of personal information revealed to social networking websites.
Self-disclosure accuracy	Willingness to provide accurate and complete personal information to social networking websites.

Appendix B: Method and Analysis

Table B1. Measurement Items

Factor	Item		Source
Self-disclosure accuracy [a]	SDA1	Please specify the extent to which you would falsify some of your personal information if it is asked for by social networking website within the next three years.	Son & Kim (2008)
	SDA2	Please specify the extent to which you would falsify some of your personal information if it would be used for big data analysis within the next three years.	
	SDA3	Please specify the extent to which you would refuse to give information to social networking website because you think it is too personal within the next three years.	
	DDV4	Please specify the extent to which you would refuse to give information if it would be used for big data analysis because you think it is too personal within the next three years.	
Self-disclosure concerns [b]	SDC1	I am more concerned about giving away sensitive information to a social networking website than about giving away sensitive information any other way.	Turow & Hennessy (2007)
	SDC2	I should have a legal right to know everything that a social networking website knows about me.	
	SDC3	My concern about outsiders learning sensitive information about me and my family has increased since I have joined a social networking website.	
	SDC4	I am nervous about social networking websites having information about me.	
Privacy concerns [c]	PrC1	I am concerned that the information I submit to a social media website could be misused.	Dinev & Hart (2004)
	PrC2	I am concerned that a person can find private information about me at a social media website.	
	PrC3	I am concerned about submitting information to a social media website, because of what others might do with it.	
	PrC4	I am concerned about submitting information to a social media website, because it could be used in a way I did not foresee.	
Trust [d]	T1	Social networking websites are safe environments in which to exchange information.	Dinev & Hart (2006)
	T2	Social networking websites are reliable environments in which to conduct business transactions.	
	T3	Social networking websites handle personal information submitted by users in a competent fashion.	
Awareness of big data [d]	ABD1	I am familiar with the term “big data.”	Developed
	ABD2	I am aware of the value of “big data” to individuals.	
	ABD3	I am aware of the value of “big data” to businesses.	
	ABD4	I understand the meaning of the term “big data.”	
Awareness of big data implications [d]	ABDI1	I am aware that my current information on social media websites could be collected.	Krasnova et al. (2009)
	ABDI2	I am aware that my current information on social media websites could be stored for years.	
	ABDI3	I am aware that my current information on social media websites could be analyzed in different ways.	
	ABDI4	I am aware that my current information on social media websites could be shared with different parties.	
Perceived control [d]	PeC1	I would only submit accurate and personal information at a social networking website if the site allowed me to control the information I volunteer.	Dinev & Hart (2004)

Table B1. Measurement Items

	PeC2	I would only provide accurate and personal information at a social networking website if the site allowed me to control the information they can use.	
	PeC3	Being able to control the personal information I provide to a social networking website is important to me.	
	PeC4	I would only provide accurate and personal information at a social networking website if their control policy is verified / monitored by a reputable third party.	
Perceived vulnerability [d]	PV1	I could be subjected to a malicious computer/ information security problems (e.g. virus, privacy, identity theft, hacking and etc.) at a social networking website.	Mohamed & Ahmad (2012)
	PV2	I feel my personal information at a social networking website could be misused.	
	PV3	I feel my personal information at a social networking website could be made available to unknown individuals or companies without my knowledge.	
	PV4	I feel my personal information at a social networking website could be made available to government agencies.	
	PV5	I feel my personal information at a social networking website could be inappropriately used.	
Self-efficacy [d]	SE1	I believe I have the ability to protect my personal information at a social networking website.	Mohamed & Ahmad (2012)
	SE2	It is easy for me to enable privacy measure features (e.g., public & private content) at a social networking website by myself.	

a: 7-point Likert scale: 1 (very unlikely) to 7 (very likely) (reversed)
b: 7-point Likert scale 1 (strongly disagree) to 7 (strongly agree)
c: 5-point Likert scale 1 (not at all concerned) to 5 (extremely concerned)
d: 5-point Likert scale 1 (strongly disagree) to 5 (strongly agree)
* SDA: self-disclosure accuracy, SDC: self-disclosure concerns, PrC: privacy concerns, T: trust, ABD: awareness of big data, ABDI: awareness of big data implications, PeC: perceived control, PV: perceived vulnerability, SE: self-efficacy.

Table B2. Confirmatory Factor Analysis Unstandardized Loadings

	Model 1	Model 2	Model 3	Model 4
Self-disclosure accuracy				
SDA1	1.000	1.000	1.000	1.000
SDA2	1.110 (0.139)	1.110 (0.155)	1.107 (0.175)	1.098 (0.130)
SDA4	0.491 (0.107)	0.490 (0.111)	0.515 (0.120)	0.490 (0.107)
Self-disclosure concerns				
SDC1	1.000	1.000	1.000	1.000
SDC3	1.118 (0.143)	1.126 (0.148)	1.191 (0.156)	1.113 (0.141)
SDC4	1.139 (0.116)	1.131 (0.116)	1.133 (0.139)	1.145 (0.115)
Privacy concerns				
PrC1	1.000	1.000	1.000	1.000
PrC3	1.159 (0.102)	1.170 (0.104)	1.135 (0.103)	1.162 (0.103)
PrC4	1.139 (0.098)	1.149 (0.100)	1.134 (0.098)	1.143 (0.099)
Trust				
T1	1.000	1.000	1.000	1.000
T2	0.965 (0.090)	0.964 (0.090)	1.013 (0.096)	0.956 (0.097)
T3	0.895 (0.105)	0.894 (0.105)	0.880 (0.112)	0.890 (0.109)
Awareness of big data				
ABD1	1.000	1.000	1.000	1.000
ABD2	0.943 (0.033)	0.943 (0.033)	0.943 (0.039)	0.943 (0.033)
ABD3	1.038 (0.030)	1.038 (0.030)	1.031 (0.034)	1.039 (0.030)
ABD4	0.991 (0.025)	0.991 (0.025)	0.995 (0.026)	0.991 (0.025)
Awareness of big data implications				
ABDI1	1.000	1.000	1.000	1.000
ABDI2	1.055 (0.116)	1.055 (0.116)	1.109 (0.138)	1.055 (0.116)
ABDI3	1.151 (0.116)	1.150 (0.116)	1.208 (0.149)	1.151 (0.116)
ABDI4	1.067 (0.116)	1.067 (0.116)	1.095 (0.149)	1.067 (0.116)
Perceived control				
PeC1	1.000	1.000	1.000	1.000
PeC2	0.920 (0.097)	0.919 (0.097)	0.935 (0.109)	0.917 (0.098)
PeC4	0.575 (0.111)	0.575 (0.111)	0.583 (0.127)	0.574 (0.110)
Perceived vulnerability				
PV1	1.000	1.000	1.000	1.000
PV2	1.395 (0.188)	1.392 (0.187)	1.479 (0.237)	1.392 (0.186)
PV3	1.270 (0.168)	1.266 (0.167)	1.334 (0.207)	1.269 (0.168)
PV4	1.107 (0.155)	1.109 (0.154)	1.193 (0.189)	1.104 (0.154)
PV5	1.291 (0.163)	1.291 (0.162)	1.406 (0.207)	1.287 (0.162)
Self-efficacy				
SE1	1.000	1.000	1.000	1.000
SE2	0.717 (0.142)	0.744 (0.149)	0.755 (0.146)	0.624 (0.142)

All loadings were significant ($p < .000$)
***** SDA: self-disclosure accuracy, SDC: self-disclosure concerns, PrC: privacy concerns, T: trust, ABD: awareness of big data, ABDI: awareness of big data implications, PeC: perceived control, PV: perceived vulnerability, SE: self-efficacy.

About the Authors

Tawfiq Alashoor is a doctoral student in Computer Information Systems at Georgia State University. He received a Master's Degree in Information Systems from Pennsylvania State University Harrisburg and a Bachelor degree in Management Information Systems from King Fahd University of Petroleum and Minerals. His main research interests are information privacy and security, social media, and cloud computing. His work has appeared in the *International Journal of Cloud Computing* and in the proceedings of several conferences including the International Conference on Information Systems, the American Conference on Information Systems, and the Hawaii International Conference on System Sciences.

Sehee Han is a doctoral candidate in Public Administration at Pennsylvania State University Harrisburg. His research interests include methodology, impact evaluation, social capital, social network services, and public health, among many others. His recent work has appeared in several journals including *Social Science & Medicine*, *Regional Studies*, *New Media & Society*, and *Reproductive Toxicology*.

Rhoda C. Joseph PhD is an Associate Professor of Information Systems at Pennsylvania State University Harrisburg. She received her doctoral degree in Information Systems, from the City University of New York. Her primary teaching areas are database management and electronic commerce. Her research is focused in the areas of E-government/M-Government, IT in Emerging Economies, and Big Data and Data Analytics. Her research has appeared in journals such as *Communications of the ACM*, *IEEE Transactions on Professional Communications*, *Government Information Quarterly*, and others. She also serves on the editorial review board for several journals.