Predicting The Disclosure of Personal Information on Social Networks: An Empirical Investigation

Thomas Buckel and Frédéric Thiesse

University of Würzburg, Chair of IS Engineering, Würzburg, Germany
{thomas.buckel,frederic.thiesse}@uni-wuerzburg.de

Abstract. The present study considers factors that motivate users of social networks to publish different types of privacy-related information to friends or even the public. In contrast to prior research, we do not limit our research scope to an individual's decision-making process (i.e., the formation of behavioral intentions) but also include actual behavior as observed among a group of real Facebook users. Our objective is to test to what extent existing theory is not only capable of explaining self-disclosure decisions but also to predict subsequent behavior. We test our model using a combination of structural equation modeling and logistic regression with questionnaire data and data collected from the Facebook platform. Our results indicate that the way self-disclosure was operationalized in prior research shows low predictive power, especially when compared to predictions based on simple questions regarding an individual's sensitivity to the disclosure of personal information.

Keywords: Social networks, privacy, self-disclosure, risk perception

1 Introduction

With more than 800 million active users, Facebook is the largest and most popular social network worldwide. More than 50 percent of its members log into the network every day in their personal profile; even the least active users interact at least once per month with the website. Around 100 billion interpersonal relations are currently maintained and more than 250 million photos are uploaded onto Facebook servers per day. Not least, the enormous success the company established is also reflected by its IPO in the range of US$ 100 billion, which poses a new record in the history of Internet-based companies [14], [36].

However, the rise of Facebook was also associated with several discussions surrounding the impact that the publication of personal profiles on the network might have on the privacy of individuals and their perceptions and valuations of privacy in general. A particular phenomenon that has attracted the interest of researchers is the 'Privacy Paradox' [34], that is, the apparent gap between the personal attitude towards the protection of privacy on the Internet and the actual behavior of social network users [3]. Facebook itself argues that the growing popularity of social networks should be interpreted as an early sign of a long-term trend towards more openness

11th International Conference on Wirtschaftsinformatik,
27th February – 01st March 2013, Leipzig, Germany

regarding personal information, which might eventually become the new social norm [15]. Indeed, in recent years some mindlessness among social network users could be observed regarding the disclosure of private information if the expected outcomes outweigh the potential dangers [28], [38].

Against this backdrop, the present study considers the factors that motivate users of social networks to publish different types of privacy-related information to friends or even the public. In recent years, a number of prior studies were presented that developed and tested models for explaining user behavior by the example of Facebook or other platforms. However, as we argue in the following, prior research was limited to the investigation of behavioral intentions, whereas actual user behavior was usually not observed. As a consequence, there remains a problematic gap in the literature with regard to the predictive power of these models, which we aim to fill. For this purpose, we build upon the model presented by Krasnova et al. (2010) and extend it by new elements related to privacy preferences and information disclosure [26]. We test our model using a combination of structural equation modeling and logistic regression using a sample of questionnaire data and data collected from the Facebook platform. Our results confirm the original results by Krasnova et al. (2010) to a large extent, but also indicate that the model's ability to predict actual behavior is rather low. This especially holds in comparison to predictions based on simple questions regarding an individual's sensitivity to the disclosure of particular types of personal information.

The remainder of the paper is organized as follows. The next section provides an overview of prior research on social networking and the disclosure of privacy-related information. Next, we develop our research model and formulate a set of testable hypotheses. The following two sections describe the data collection process and the results of the hypothesis tests. The paper closes with a discussion of our findings, implications for practice and theory, and limitations.

2 Related Work

A number of prior studies can be found in the existing body of literature, which deals with the apparent gap between privacy preferences and disclosure of privacy-related information on the Internet. Berendt et al. (2005) presented a study with 171 respondents, which compared the actual behavior of internet users to previous statements on various privacy aspects. The authors reported that the participants revealed significantly more information on the net than they were willing to share according to the survey. They conjecture that expected benefits from the disclosure of information might be the predominant reason for this phenomenon [3].

Acquisti et al. (2006) found that users are increasingly aware of privacy risks associated with social media. In their study more than 500 Facebook members were surveyed and 196 profiles were analyzed. The objective of the study was to compare the congruence of survey responses with the actual profiles. Among others, the respondents were asked whether a particular type of information had been published within their Facebook profile or not. The results indicate an 80% match. Notwithstanding

some confusion with regard to the the complexity of the Facebook privacy settings, most of the respondents were aware of their level of self-disclosure [1].

Dwyer et al. (2007) developed a model, which explains information disclosure and the emergence of relationships on Facebook and Myspace by the users' trust in the platform and its members. According to the answers given by the study participants, the more information is shared the more users trust the other parties. Moreover, the authors tried to predict this behavior by the factor 'Internet Privacy Concern', but the empirical results do not indicate any significant influence [13].

Krishnamurthy & Wills (2008) found that 55% to 90% of social network members make their profiles publicly available. At least 80% provide confirmed contacts with complete access to their personal profile. In addition, their results indicate a negative relation between the size of the network and the use of functions for limiting the public visibility of profiles [27].

Christofides et al. (2009) conducted a survey and came to the counterintuitive conclusion that the willingness to publish personal information may not be negatively correlated with the user's perceived control of their data. Both aspects were defined as independent behaviors, with the general tendency to disclose, self-esteem and trust in the platform being the strongest predictors for information control [10].

Another study by Debatin et al. (2009) confirmed the conjecture by Youn (2009) that the expected benefit of being a Facebook user outweighs the perceived risks [11], [38]. In the same year, Bonneau et al. (2009) studied the manifold options for extracting profile data from other users by circumventing the network's security mechanisms [4]. The security issues associated with third-party applications, which are integrated into Facebook, were also confirmed by Felt (2008) [16]. Data protection problems in social networks in general were discussed by Fung et al. (2010) [18].

The study by Hoy & Milne (2010) set the focus on the differences between men and women in the context of privacy risks on the Facebook platform. The authors found that female users are slightly more sensitive to privacy risks [21].

Krasnova et al. (2010) investigated the motivation for disclosing personal information. In their model, self-disclosure is traced back to the negative influence of perceived risk as well as the positive influence of enjoyment, convenience, and the desire for self-presentation and relationship building. The empirical results show strong support for their model with the independent variables explaining about half of the variance in the self-disclosure variable [26].

In sum it can be said that prior research on self-disclosure in social networks has contributed substantially to our understanding of the decision-making process that eventually leads to people publishing privacy-related information. However, as with many other studies following the behaviorist paradigm, the scope of the empirical analysis is all too often limited to the investigation of behavioral intentions without observing actual behavior. Although behavioral intention is known to be a strong predictor of behavior, there is an evident gap in the literature regarding the ability of prior research models to predict actual behavior of social network users.

3 Research Model

The research model underlying the present study is depicted in Figure 1. Following the model proposed by Krasnova et al. (2010), we posit that perceived risks associated with the disclosure of personal information exert a negative influence on self-disclosure of an individual. We assume that risk perception is to a large extent determined by perceived control, trust in the social network operator, and trust in other network members. Self-disclosure, again, is assumed to be influenced by convenience, enjoyment, self-presentation, and the opportunity to build relations with others [26]. We extend this original model by an additional construct, which reflects an individual's preferences regarding the disclosure of specific types of privacy-related information. We hypothesize that this additional construct acts as mediator between risk perception and self-disclosure. Furthermore, we assume that the three dependent variables in the model may allow for predicting actual disclosure of personal information on the social network.

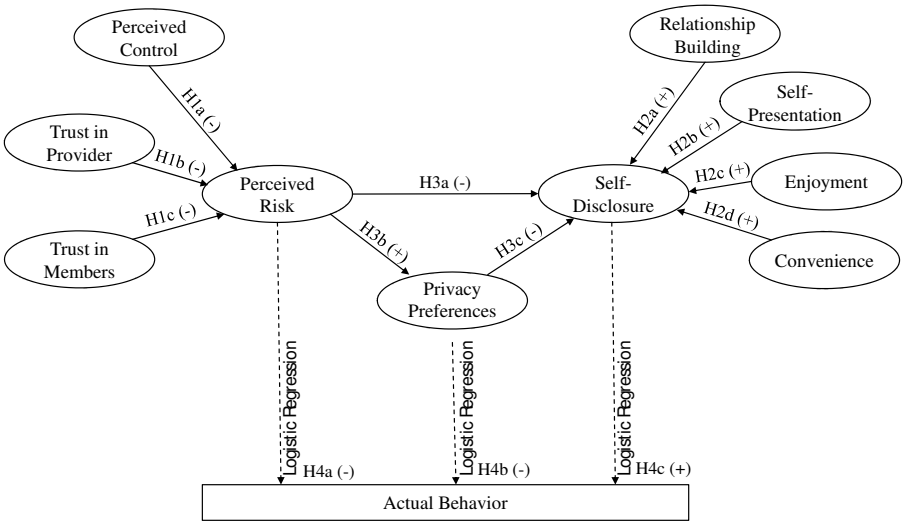

Fig. 1. Structural research model and hypotheses

Hypotheses H1a/b/c and H2a/b/c/d in our model were adapted from Krasnova et al. (2010). Due to space considerations, we refer the reader to this prior study for the underlying theoretical rationale. In the following, we limit ourselves to a mere enumeration of the hypotheses:

H1a: Perceived control exerts a negative influence on the perceived risk associated with the use of a social network platform.

H1b: Trust in other members of the network exerts a negative influence on the perceived risk associated with the use of a social network platform.

H1c: Trust in the operator of the social network exerts a negative influence on the perceived risk associated with the use of a social network platform.

H2a: Users' beliefs regarding relationship-building opportunities exert a positive influence on their self-disclosure on a social network platform.

H2b: Users' perceived benefits of self-presentation opportunities exert a positive influence on their self-disclosure on a social network platform.

H2c: Users' perceived enjoyment of platform use exerts a positive influence on their self-disclosure on a social network platform.

H2d: Users' beliefs regarding a network's ability to aid them in conveniently maintaining relationships exerts a positive influence on their self-disclosure on a social network platform.

The model by Krasnova et al. (2010) also posits a negative causal relation between the perceived risk and self-disclosure based on prior research, among others, by Malhotra (2004) [26], [29]:

H3a: Users' perceived risk associated with the use of the network exerts a negative influence on their intention to disclose privacy-related information.

An extension that we make to this original model refers to the modeling of risk beliefs and privacy preferences. While Krasnova et al. (2010) do not distinguish between the two and posit a direct effect of perceived risk on self-disclosure, we conjecture a more complex relation including a mediating variable. This view is supported by prior studies, which indicate that general risk beliefs materialize in the form of more specific concerns and/or preferences regarding the disclosure of personal information [12], [26]. These concerns/preferences may then have an impact on the intention to disclose information on the social network [12], [25], [38]. Consequently, we hypothesize that the relation between risk beliefs and self-disclosure might be mediated by an individual's privacy preferences regarding specific types of privacy-related information:

H3b: Perceived risks associated with the use of the network exert a positive influence on the formation of an individual's preferences regarding the disclosure of privacy-related information.

H3c: Privacy preferences regarding the disclosure of privacy-related information exert a negative influence on self-disclosure on the social network.

A second extension of the original model refers to our objective of investigating not only behavioral intentions but rather to predict actual behavior. Prior research shows inconsistencies between an individual's statements on self-disclosure and the information provided by the very same person in social media [1], [13]. Norberg et al. (2007) conclude that in the context of privacy research generalized constructs might be too coarse-grained to be applicable to the prediction of actual behavior [30]. We hence aim to test whether the two dependent variables from the original model as well as the new introduced privacy preferences pose predictors of actual behavior. The underlying rationale is that general constructs and the corresponding measurement scales are inferior to more specific questions regarding individual preferences when it comes to the prediction of self-disclosure behavior. If our assumption holds true, this might explain the apparent privacy paradox observed in prior research by a measurement issue rather than a theoretical gap.

H4a: Users' perceived risk associated with the use of the network exerts a negative influence on actual self-disclosure behavior.

H4b: Users' privacy preferences with regard to the disclosure of privacy-related information exert a negative influence on actual self-disclosure behavior.

H4c: Users' intention to self-disclose on the network exerts a positive influence on actual self-disclosure behavior.

Table 1. Questionnaire and construct operationalization

Construct	Items
Perceived Control (PC)	1. I feel in control over the information I provide on Facebook 2. Privacy settings allow me to have full control over the information I provide 3. I feel in control of who can view my information on Facebook
Trust in Provider (TP)	1. Facebook is open and receptive to the needs of its members 2. Facebook makes good-faith efforts to address most member concerns 3. Facebook is also interested in the well-being of its members, not just its own 4. Facebook is honest in its dealings with me 5. Facebook keeps its commitments to its members 6. Facebook is trustworthy
Trust in Members (TM)	1. Other members will do their best to help me 2. Other members care about the well-being of others 3. Other members are open and receptive to the needs of each other 4. Other members are honest in dealing with each other 5. Other members keep their promises 6. Other members are trustworthy
Perceived Risk (RISK)	1. Overall, I see no real threat to my privacy due to my presence on Facebook 2. I fear that something unpleasant can happen to me due to my presence on Facebook 3. Overall, I find it risky to publish my personal information on Facebook 4. Please rate your overall perception of privacy risk involved when using Facebook
Relationship Building (RB)	1. Through Facebook I get connected to new people who share my interests 2. Facebook helps me to expand my network 3. I get to know new people through Facebook.
Self-Presentation (SPR)	1. I try to make a good impression on others on Facebook 2. I try to present myself in a favourable way on Facebook 3. Facebook helps me to present my best sides to others
Enjoyment (EN)	1. When I am bored I often login to Facebook 2. I find Facebook entertaining 3. I spend enjoyable and relaxing time on Facebook
Convenience (CON)	1. Facebook is convenient to inform all my friends about my ongoing activities 2. Facebook allows me to save time when I want to share new stuff with my friends 3. I find Facebook efficient in sharing information with my friends
Self-Disclosure (SD)	1. I have a comprehensive profile on Facebook 2. I find time to keep my Facebook-profile up-to-date 3. I keep my friends updated about what is going on in my life through Facebook 4. When I have something to say, I like to share it on Facebook
Privacy Preferences (PP)	1.- 21. How crucial do you rate the availability of your personal information (‘Name”, ‘Home address”, ‘Hometown”, ‘Date of birth”, …) on the internet?

4 Data Collection

Constructs from the research model were operationalized using multi-item measurement scales with at least three items per construct. All questions were measured by 7-

point Likert scales (see Table 1). The corresponding scales were adapted from Krasnova et al. (2010) with the exception of the privacy preferences construct, which was modeled as a formative construct based on the different types of personal information included in a Facebook profile. The questionnaire was discussed with external experts and tested before the actual data collection with a group of students in order to ensure the comprehensibility of our questions. The questionnaire was prepared in the form of a PDF file, which avoids some issues with data quality associated with online surveys and allows the respondents to interrupt the process at any time and to continue later.

The data were collected from undergraduate and graduate students of business administration, economics and MIS. In total, 650 students were contacted. In addition to the questionnaire, we also asked respondents to accept 'friend requests' sent from our institute's Facebook profile. The latter step was necessary in order to access not only the respondents' public information but also information that is restricted to friends. In total, 182 students accepted our friend requests, among which 105 filled out the questionnaire. Five questionnaires were incomplete and had to be excluded from further analysis. As a result, we received 100 usable responses, which equals a response rate of 15.4%. The respondents' age was in the range of 17 to 30 years. On average, they were members of Facebook since 2009 and spend about 1.7 hours per day in the network. Genders were almost equally distributed with 51 male and 49 female respondents. The number of Facebook friends per respondent lies in the range of 150 to 300 persons. Most of the respondents were also members of other social networks.

5 Data Analysis

5.1 Structural Model Test

In a first step, we tested our research model excluding actual disclosure behavior using structural equation modeling techniques. For this reason, we followed the original procedure applied by Krasnova et al. (2010), who used the Partial Least Squares (PLS) method with Bootstrapping and Blindfolding [35]. A second reason for the use of PLS was the fact that our model included a formative construct, which cannot be modeled using covariance-based techniques [8]. Furthermore PLS is preferred instead of LISREL as the method is better suited for optimizing predictive power [6-7].

Validity and reliability of our scales were verified in a factor analysis. Bartlett's test of sphericity was accepted in all cases, which already indicates independent constructs. Items with factor loadings smaller than 0.5 were removed [23]. The number of items per factor was nevertheless three or greater for all factors. We continued with verifying the fit of the internal structure of our model by examining Cronbach's Alpha, composite reliability and the average variance extracted (AVE) per factor. The only exception was the 'Trust in Members' factor with an AVE of 0.46. Since the value is only slightly below the threshold and the factor did not contribute substantially to the overall explanatory power of the model by Krasnova et al. (2010), we decided not to exclude it. All factors surpass the widely accepted thresholds [2], [31]. Dis-

criminant validity was ensured by verifying that the square root of the respective AVE values was also larger than the correlations between the construct's validity [17].

We then calculated model fit indicators for the structural model. The PLS method does not offer a set of global fit indicators in the same way as covariance-based techniques. The fit of the model to the data can hence only be judged by R^2 and Q^2 values for the dependent variables. In our case, the R^2 values are greater than 0.2 for all relevant factors, which indicates substantial explanatory power [8]. Furthermore, we examined the Stone-Geisser criterion (Q^2), which indicates the predictive relevance [19], [33]. We consider both the cross-validated redundancy (Q^2 cvr) as well as the cross-validated communality (Q^2 cvc) [8-9], [24], [35]. Both types of Q^2 values are greater than 0 for all relevant factors. An overview of all considered factor and model criteria is given in Table 2.

The results of the actual hypothesis tests are given in Figure 2. In the following, we not only provide path coefficients and confidence levels, but also the f^2 statistic, which indicates the effect size [20], [32]. The test results show that PC exerts a significant negative influence on RISK (beta = -0.367, $f^2 = 0.15$, $p < 0.001$). In contrast, the influence of TP (beta = -0.186, $f^2 = 0.04$, $p > 0.05$) and TM (beta = -0.108, $f^2 = 0.01$, $p > 0.05$) is not significant. RB exerts a significant positive influence on SD (beta = 0.306, $f^2 = 0.14$, $p < 0.001$); the same holds for the influence of SPR (beta = 0.221, $f^2 = 0.08$, $p < 0.01$) and EN (beta = 0.229, $f^2 = 0.06$, $p < 0.05$) on SD. In contrast, the influence of CON on SD (beta = 0.070, $f^2 = 0.00$, $p > 0.05$) remains insignificant. In sum, our results are partly consistent with those presented by Krasnova et al. (2010), who were able to confirm a significant influence of TP on RISK. Furthermore, the results by Krasnova et al. (2010) indicate a significant influence of SP on SD, whereas the influence of CON was insignificant [26]. Our results show the opposite. We can thus comfirm H1a and H2a/b/c, whereas H1b/c and H2d must be rejected.

Table 2. Validity, reliability, and model fit indicators

Factor	Type	Alpha	CR	AVE	R^2	Q^2 cvr	Q^2 cvc
Threshold		***≥ 0.7***	***≥ 0.7***	***≥ 0.5***	***≥ 0.2***	***≥ 0***	***≥ 0***
Perceived Control (PC)	R	0.88	0.93	0.81	-	-	0.59
Trust in Provider (TP)	R	0.86	0.88	0.55	-	-	0.37
Trust in Members (TM)	R	0.87	0.83	0.46	-	-	0.14
Perceived Risk (RISK)	R	0.77	0.85	0.59	0.23	0.12	0.31
Relationship Building (RB)	R	0.77	0.86	0.68	-	-	0.36
Self-Presentation (SPR)	R	0.89	0.93	0.82	-	-	0.61
Enjoyment (EN)	R	0.76	0.85	0.67	-	-	0.36
Convenience (CON)	R	0.72	0.82	0.61	-	-	0.24
Self-Disclosure (SD)	R	0.86	0.90	0.70	0.57	0.35	0.50
Privacy Preferences (PP)	F	-	-	-	0.26	0.04	0.14

Notes: R = reflective; F = formative

With regard to the interrelations between the dependent variables, we observed a significant positive influence of RISK on PP (beta = 0.541, $f^2 = 0.37$, $p < 0.001$). Furthermore, the negative influence of PP on SD (beta = -0.294, $f^2 = 0.16$, $p < 0.05$) was significant, too. In contrast, the results did not indicate a significant relation between

RISK and SD (beta = -0.041, $f^2 = 0.00$, $p > 0.05$), which can be explained by the inclusion of PP as mediator between the two factors. We can thus confirm hypotheses H3b/c, whereas H3a must be rejected. The R^2 value for SD shows that the model is able to explain 57.1% of the observed variance in this pivotal factor.

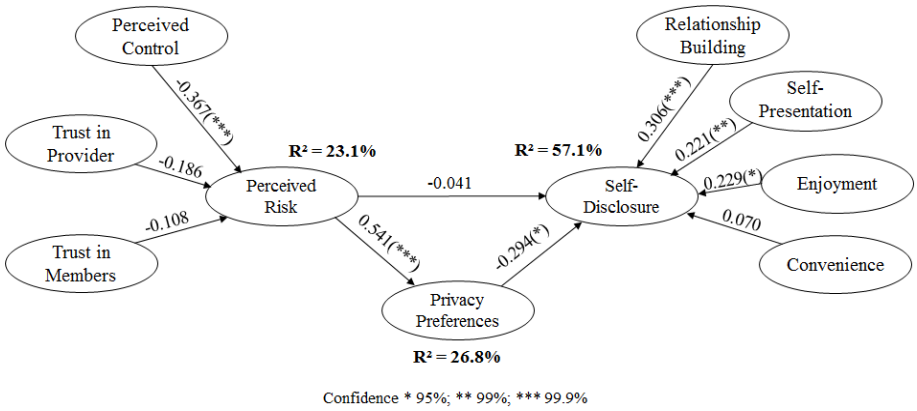

Fig. 2. Results of the Structure Model Test

5.2 Logistic Regression

In a second analysis step, we investigated to what extent the model is not only able to explain behavioral intentions (i.e., the variance observed in the SD construct), but also actual behavior. It is important to note that the explanatory power of the model measured by the R^2 value for SD should not be mistaken for the model's predictive power. For this reason, we considered actual behavior in the form of personal information published in an individual's Facebook profile. For each information type (e.g., name, address, religion), we investigated whether the dependent variables from the structural model act as predictors. Since actual behavior is measured by a set of binary variables, we use the logistic regression method for analysis. We consider both profile information that is publicly available as well as information that is restricted to friends. The occurrence of the 20 different information types on the Facebook profile pages of our survey respondents is summarized in Table 3. Note that 'Name' and 'Profile picture' are always publicly available. We considered only information types, which were included in at least 10% (average of public and friends profile) of the respondents' profiles. As an exceptional case, the mobile number was included though this information type did not reach the 10% threshold. The 'restriction ratio' denotes the proportion of people, who restricted a particular type of information to their friends.

Table 3. Occurrence of Facebook profile information among the survey respondents

Information type	Visibility	%	Restriction ratio	Information type	Visibility	%	Restriction ratio
Name	Friends / Public	74%	-	Current school	Friends	74%	43%
					Public	42%	
Home address	Friends	64%	47%	Former school	Friends	61%	51%
	Public	34%			Public	30%	
Home-town	Friends	53%	43%	Internships	Friends	2%	100%
	Public	30%			Public	0%	
Date of birth	Friends	83%	95%	Profile picture	Friends / Public	92%	-
	Public	4%					
E-Mail	Friends	70%	97%	Other pictures	Friends	84%	89%
	Public	2%			Public	9%	
Mobile number	Friends	4%	100%	Interests	Friends	86%	59%
	Public	0%			Public	35%	
Phone number	Friends	0%	-	Sports	Friends	39%	64%
	Public	0%			Public	14%	
Home-page	Friends	5%	80%	Politics	Friends	5%	80%
	Public	1%			Public	1%	
Marital status	Friends	60%	88%	Religion	Friends	9%	100%
	Public	7%			Public	0%	
Current profession	Friends	12%	42%	Relationships	Friends	37%	86%
	Public	7%			Public	5%	
Former profession	Friends	9%	67%	Avg. visible for friends: 42% Avg. visible for public: 19%			Avg. ratio: 74%
	Public	3%					

We developed separate logistic regression models for each type of profile information (= dependent variable) and the RISK, PP, and SD constructs from our structural model (= independent variables). In the case of the PP construct, we considered each item separately as predictor to the corresponding information type. In total, we investigated the prediction of 13 out of 21 information types among the restricted profiles and 10 out 21 among the public profiles. The results of this analysis step are given in Table 4. We provide information on Nagelkerke's R^2, the regression coefficients, and the corresponding significance value. Based on the R^2 value, we also determine which predictor works best for the respective information type. Our results show that the simple question for an individual's privacy preference works best as predictor. In contrast, the more general SD factor shows much lower R^2 values and, in some cases, exerts no statistically significant influence at all. The same holds for the RISK factor, which is the best predictor in only one case. In total, the logistic regression analysis indicates that the structural model, which performs reasonably well with regard to the explanation of behavioral intentions, is limited when it comes to the prediction of actual behavior. We therefore concluded that H4b can be confirmed, while H4a/c should rather be rejected.

Table 4. Logistic Regression Analysis Results

Independent Variable	Dependent Variable	Profile	Nagelkerkes R^2	Regression Coefficient	Sig.	Best Predictor
PP_Name	Name	F/P	0.289	-0.616 (***)	0	
RISK		F/P	0.067	-0.576 (*)	0.037	PP
SD		F/P	0.078	0.633 (*)	0.029	
PP_HomeAddress	Home address	F	0.122	-0.354 (***)	0.004	
RISK			0	-0.037	0.873	PP
SD			0.061	0.505 (*)	0.042	
PP_HomeAddress		P	0.036	-0.185	0.109	
RISK			0	0.005	0.983	SD
SD			0.085	0.577 (*)	0.015	
PP_Hometown	Hometown	F	0.159	-0.41 (***)	0.001	
RISK			0	0.021	0.926	PP
SD			0.137	0.768 (*)	0.002	
PP_Hometown		P	0.043	-0.205	0.083	
RISK			0.004	-0.126	0.604	SD
SD			0.077	0.557 (*)	0.02	
PP_DateOfBirth	Date of birth	F	0.163	-0.523 (**)	0.003	
RISK			0.142	-0.986 (**)	0.007	PP
SD			0.06	0.611	0.075	
PP_EMail	E-Mail	F	0.117	-0.448 (**)	0.007	
RISK			0.092	-0.666 (*)	0.013	PP
SD			0.001	0.072	0.761	
PP_Handy	Mobile number	F	0.185	-0.609 (***)	0.005	
RISK			0.008	-0.251	0.592	PP
SD			0.056	0.616	0.149	
PP_MaritalStatus	Marital status	P	0.097	-0.46 (*)	0.05	
RISK			0.053	-0.647	0.157	PP
SD			0	0.04	0.925	
PP_CurSchool	Current school	F	0.241	-0.605 (***)	0	
RISK			0.147	-0.905 (**)	0.003	PP
SD			0.084	0.661 (*)	0.024	
PP_CurSchool		P	0.124	-0.409 (**)	0.004	
RISK			0.012	-0.218	0.34	PP
SD			0.021	0.276	0.211	
PP_ForSchool	Former school	F	0.074	-0.294 (*)	0.021	
RISK			0.012	-0.215	0.355	SD
SD			0.184	0.969 (***)	0.001	
PP_ForSchool		P	0.047	-0.255	0.079	
RISK			0.004	-0.136	0.576	SD
SD			0.059	0.481 (*)	0.043	
PP_ProfPicture	Profile picture	F/P	0.138	-0.595 (*)	0.027	
RISK			0.058	-0.696	0.132	SD
SD			0.232	1,988 (*)	0.013	
PP_OthPictures	Other pictures	F	0.087	-0.496 (*)	0.046	
RISK			0.063	-0.623	0.063	PP
SD			0.069	0.675	0.061	
PP_OthPictures		P	0.134	-0.649 (*)	0.016	
RISK			0.005	-0.182	0.64	PP
SD			0.097	0.762 (*)	0.037	
PP_Interests	Interests	F	0.062	-0.338	0.073	
RISK			0.100	-0.836 (*)	0.026	RISK
SD			0.007	0.199	0.541	
PP_Interests		P	0.130	-0.409 (**)	0.003	
RISK			0.001	-0.076	0.747	PP
SD			0.015	0.234	0.297	
PP_Sports	Sports	F	0.077	-0.303 (*)	0.019	
RISK			0.001	0.065	0.778	PP
SD			0.067	0.505 (*)	0.028	
PP_Sports		P	0.092	-0.428 (*)	0.034	
RISK			0.007	-0.199	0.536	PP
SD			0.068	0.588 (*)	0.051	
PP_Relationships	Relationships	F	0.189	-0.477 (***)	0	
RISK			0.006	0.151	0.517	PP
SD			0.057	0.464 (*)	0.043	

6 Implications for Theory and Practice

In this section, we compare the results from the structural model test to the results by Krasnova et al. (2010) and discuss our original results, which go beyond this piece of prior resarch. With regard to the interrelations between factors of the structural model, we found that neither the influence of trust in other members nor trust in the network operator showed a statistically significant influence on the users' risk perception. This phenomenon might be explained by the increasing understanding on the part users that the use of Facebook poses a privacy risk and that the platform operator makes money from the processing of personal information. The only factor that could be confirmed to exert a significant influence is perceived control. This might be attributed to the growing knowledge among users regarding the practical application of different privacy settings [5].

Similar to Krasnova et al. (2010), we were able to confirm significant positive influences of the ability to maintain relationships and enjoyment from Facebook usage on the intention to self-disclose. In contrast to this prior study, we also observed that the opportunity for self-presentation has a positive impact on self-disclosure, too. On the other hand, we could not confirm an influence of convenience, which had been identified by Krasnova et al. (2010) as a significant factor. Reasons for these different results might be that the ease of use of Facebook has become a standard among social network websites and that users' have become used to it. Moreover, the increasing awareness of the benefits of self-presentation might explain the impact of the users's interest in self-presentation on self-disclosure [5], [28]. Furthermore, we were able to confirm the relation between risk perception and self-disclosure. However, we extended the research model by an additional mediator variable, which reflects the users' preferences regarding the publication of private information. Our results indicate that privacy preferences are to a large extent shaped by the perceived risk and determine an individual's willingness to self-disclose. In contrast to that, the direct effect between perceived risk and self-disclosure did not turn out to be significant. In addition, we saw that our refined model shows a higher explanatory power with $R^2 = 0.571$ for the self-disclosure variable compared to 0.472 in the study by Krasnova et al. (2010).

In order to test the predictive power of the dependent variables with regard to actual behavior, we conducted a logistic regression analysis for each of the personal information types, which are part of any Facebook profile. The results show that the predictive power of the general factors from prior research is low and particularly inferior to simple questions regarding an individual's privacy preferences. This finding reflects a problematic assumption made in many behaviorist studies, which assume that actual behavior is in virtually any case determined by behavioral intentions and may hence be excluded from the respective investigation. In contrast, our results highlight the need for collecting data on actual behavior as well, even though the costs for acquiring such data are usually much higher than conducting a survey alone.

From a practitioner's perspective, the study also allows for a number of implications. Empirical studies of self-disclosure behavior are important means for judging the relevance of social network design decisions. Our results confirm some of the factors identified by prior research. We have seen that self-presentation and enjoy-

ment are important influence factors, which must be supported by any social network platform. We have also seen that perceived control exerts a significant influence on the perception of privacy risks. Social network operators should implement mechanisms that allow for fine-grained control of information disclosure. However, we could also show that investigations of general intentions and attitudes are not sufficient for analyzing the future behavior of users and that more specific measurement instruments are needed in order to capture their real privacy preferences.

7 Summary and Outlook

The purpose of the present study was to address a common weakness in many behaviorist studies considering human decision-making processes. Despite the fact that behavioral intentions are known to be strong predicitors of actual behavior, researchers should not take this relation for granted and include data on their respondents' behavior in their analysis. We considered the specific example of self-disclosure on social networks, which has been the subject of a substantial number of prior studies. Based on our review of the literature, we found that the majority of existing studies relied only on survey data, whereas actual self-disclosure was usually ignored.

In order to fill this gap in the literature and to evaluate the predictive power of existing models, we extended and tested the research model by Krasnova et al. (2010), which showed high explanatory power with regard to the self-disclosure variable. We complemented the data set collected via a survey among 100 Facebook users by data on the same individuals' personal profiles on the social network. We evaluated the predictive power of the self-disclosure variable and compared it to a set of simple questions related to the respondents' privacy preferences. The results show that the factor measuring behavioral intentions is a rather weak predictor for actual behavior. This result could be interpreted as a sign of the so-called 'Privacy Paradox' phenomenon. However, our results also show the stated privacy preferences allow for much better predictions of the same data. This leads us to the conclusion that the Privacy Paradox may to some extent be attributed to a measurement issue in contrast to a more fundamental flaw in our theoretical understanding of self-disclosure.

Limitations of the study could be seen in the relatively small data sample, which we nevertheless consider sufficient for an analysis using PLS and logistic regression. Second, it should be noted that the respondents' survey answers and profiles might have been influenced by their participation in the study. The latter is a common issue with any study that depends on respondents who are aware of being observed. Further research will be necessary in two regards. On the one hand, researchers should strive to confirm our results using different and larger data samples. On the other hand, further research will be necessary in order to develop refined models and measurement scales to improve the explanatory and predictive power, for example, by introducing constructs such as 'social norm' [22], [37] into the respective models.

References

1. Acquisti, A., Gross, R.: Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In: Proceedings from Privacy Enhancing Technologies Workshop. Cambridge, UK (2006)
2. Barker, C., Pistrang, N., Elliott, R.: Reserach methods in clinical and counseling psychology. John Wiley, Chichester (1994)
3. Berendt, B., Günther, O., Spiekermann, S.: Privacy in E-Commerce: Stated Preferences vs. Actual Behavior. Communication of the ACM 48 (3), 101-106 (2005)
4. Bonneau, J., Anderson, J., Danezis, G.: Prying Data Out of a Social Network. In: Proceedings of the International Conference on Advances in Social Network Analysis and Mining, pp. 249-254, Washington, DC, USA (2009)
5. Boyd, D.: Facebook's Privacy Trainwreck: Exposure, Invasion, and Social Convergence. The International Journal of Research into New Media Technologies, (14) 1, 13-20 (2008)
6. Chin, W.W., Marcolin, B.L., Newsted, P.R.: A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic–Mail Emotion/Adoption Study. Information Systems Research 14 (2), 189-217 (2003)
7. Chin, W.W., Newsted, P.: Structural Equation Modeling Analysis with small Samples using PLS. In: Hoyle, R.H.: Statistical Methods for small sample research. Sage Publications, Thousand Oaks et al. (1999)
8. Chin, W.W.: Issues and Opinion on Structural Equation Modeling. MIS Quarterly 22 (1), vii-xvi (1998)
9. Chin, W.W.: The partial least squares approach to structural equation modeling. In: Marcoulides, G.A. (eds.): Modern methods for business research. Lawrence Erlbaum, Mahwah (1998)
10. Christofides, E., Muise, A., Desmarais, S.: Information Disclosure and Control on Facebook: Are They Two Sides of the Same Coin or Two Different Processes? Cyberpsychology & Behavior 12 (3), 341-345 (2009)
11. Debatin, B., Lovejoy, J., Horn, A., Hughes, B.N.: Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. Journal of Computer-Mediated Communication 15 (1), 83-108 (2009)
12. Dinev, T., Hart, P.: An Extended Privacy Calculus Model for E-Commerce Transactions. Information Systems Research 17 (1), 61-80 (2006)
13. Dwyer, C., Hiltz, S.R., Passerini, K.: Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In: Proceedings of the Thirteenth Americas Conference on Information Systems, Keystone (CO), Paper 339 (2007)
14. Facebook, http://www.facebook.com
15. Farber, A.: Facebook's Zuckerberg calls online openness the 'social norm.' New Media Age, 1/14/2010, 05 (2010)
16. Felt, A., Evans, D.: Privacy Protection for Social Networking Platforms. In: Proceedings of W2SP 2008: Web 2.0 Security and Privacy, pp 1-8, Oakland (CA) (2008)
17. Fornell, C., Larcker, D.F.: Evaluating Structural Equation Models with unobservable variables and measurment error. Journal of Marketing Research 18 (2), 39-50 (1981)
18. Fung, B.C.M., Wang, K., Chen, R., Yu, P.S.: Privacy-preserving data publishing: A survey of recent developments. ACM Computing Surveys (CSUR) 42 (4), Article 14 (2010)
19. Geisser, S.: A Predictive Approach to the Random Effects Model. Biometrica 61, 101-107 (1974)

20. Henseler, J., Fassott, G., Dijkstra, T.K., Wilson, B.: Analysing quadratic effects of formative constructs by means of variance-based structural equation modelling. European Journal of Information Systems 21 (1), 99-112 (2012)
21. Hoy, M.G., Milne, G.: Gender Differences in Privacy-Related Measures for Young Adult Facebook Users. Journal of Interactive Advertising 10 (2), 28-45 (2010)
22. Johnston, A.C., Warketin, M.: Fear Appeals and Information Security Behaviors: An Empirical Study. MIS Quaterly 34 (3), 549-566 (2010)
23. Kaiser, H.F.: An index of factorial simplicity. Psychometrika 39 (1), 31-36 (1974)
24. Karim, J.: Emotional Labor and Psychological Distress: Testing the Mediatory Role of Work-Family Conflict. European Journal of Social Sciences 11 (4), 584-598 (2009)
25. Krasnova, H., Kolesnikova, E., Günther, O.: It Won't happen to me!: Self-Disclosure in Online Social Networks. In: Proceedings of the Fifteenth Americas Conference on Information Systems, Paper 343, San Francisco (2009)
26. Krasnova, H., Spiekermann, S., Koroleva, K., Hildebrand, T.: Online social networks: why we disclose. Journal of Information Technology 25, 109–125 (2010)
27. Krishnamurthy, B., Wills, C.E.: Characterizing privacy in online social networks. In: Proceedings of the Workshop on Online Social Networks in conjunction with ACM SIGCOMM Conference, pp. 37-42, Seattle (WA) (2008)
28. Lucas, M.M., Borisov, N.: FlyByNight: mitigating the privacy risks of social networking. In: Proceedings of the 7th ACM workshop on Privacy in the electronic society, pp. 1-8, Alexandria (VA) (2010)
29. Malhorta, N.K., Kim, S.S., Agarwal, J.: Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. Information Systems Research 15 (4), 336-355 (2004)
30. Norberg, P.A., Horne, D.R., Horne, D.A.: The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. The journal of consumer affairs 41 (1), 100-126 (2007)
31. Nunnally, J.C., Bernstein, I.H.: Psychometric Theory. Mc Graw-Hill, New York (1994)
32. Ringle, C.M., Sarstedt, M., Straub D.W.: A Critical Look at the Use of PLS-SEM. MIS Quarterly 36 (1), iii-xiv (2012)
33. Stone, M.: Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Society. Series B (Methodological) 36 (2), 111-147 (1974)
34. Stutzman, F., Vitak, J., Ellison, N., Gray, R., Lampe, C.: Privacy in Interaction: Exploring Disclosure and Social Capital in Facebook. In: International Conference on Weblogs and Social Media (ICWSM '12), Dublin, IE (2012)
35. Tenenhaus, M., Vinzi, V.E., Chatelin, Y., Lauro, C.: PLS path modeling. Computational Statistics and Data Analysis 48 (1), 159-205 (2005)
36. Townsend, A.: This Is Your Life (According to Your New Timeline). Time International (South Pacific Edition) 179 (6), 24-27 (2012)
37. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. MIS Quaterly 27 (3), 425-478 (2003)
38. Youn, S.: Determinants of Online Privacy Concern and Its Influence on Privacy Protection Behaviors Among Young Adolescents. Journal of Consumer Affairs 43 (3), 389-418 (2009)