Human–Computer Interaction

ISSN: 0737-0024 (Print) 1532-7051 (Online) Journal homepage: http://www.tandfonline.com/loi/hhci20

Not Even Past: Information Aging and Temporal Privacy in Online Social Networks

Oshrat Ayalon & Eran Toch

To cite this article: Oshrat Ayalon & Eran Toch (2017) Not Even Past: Information Aging and Temporal Privacy in Online Social Networks, Human–Computer Interaction, 32:2, 73-102, DOI: 10.1080/07370024.2016.1203791

To link to this article: http://dx.doi.org/10.1080/07370024.2016.1203791

Accepted author version posted online: 08 Jul 2016.
Published online: 08 Jul 2016.

HUMAN–COMPUTER INTERACTION, 2017, Volume 32, pp. 73–102

ISSN: 0737-0024 print / 1532-7051 online
DOI: 10.1080/07370024.2016.1203791

Not Even Past: Information Aging and Temporal Privacy in Online Social Networks

Oshrat Ayalon and Eran Toch
Tel Aviv University, Israel

Online social networks (OSNs) make information accessible for unlimited periods and provide easy access to past information by arranging information in time lines or by providing sophisticated search mechanisms. Despite increased concerns over the privacy threat that is posed by digital memory, there is little knowledge about retrospective privacy: the extent to which the age of the exposed information affects sharing preferences. In this article, we investigate how information aging impacts users' sharing preferences on Facebook. Our findings are based on a between-subjects experiment ($n = 272$), in which we measured the impact of time since first publishing an OSN post on its sharing preferences. Our results quantify how willingness to share is lower for older Facebook posts and show that older posts have lower relevancy to the user's social network and are less representative of the user's identity. We show that changes in the user's social circles, the occurrence of significant life changes and a user's young age are correlated with a further decrease in the willingness to keep sharing past information. We discuss our findings by juxtaposing digital memory theories and privacy theories and suggest a vision for mechanisms that can help users manage longitudinal privacy.

Oshrat Ayalon (oshratra@post.tau.ac.il, www.oshratayalon.com) is an Information Systems Scientist with an interest in human-computer interaction and usable privacy and security; she is a PhD student in the Department of Industrial Engineering of Tel Aviv University. **Eran Toch** (erant@post.tau.ac.il, http://toch.tau.ac.il/) is an Information Systems Scientist with an interest in human–computer interaction, privacy, and artificial intelligence; he is a Senior Lecturer (equivalent to an Assistant Professor) at the Department of Industrial Engineering of Tel Aviv University.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/HHCI.

CONTENTS

1. TIME AND INFORMATION SHARING IN ONLINE SOCIAL NETWORKS

> "The past is never dead. It's not even past." — *Requiem for a Nun* (William Faulkner, 1951)

Online social networks (OSNs) make personal information accessible for long periods of time. Information that is uploaded to OSNs is not deleted by default (Gross & Acquisti, 2005; Rosenblum, 2007), which results in increasingly large amounts of personal data that are easily accessible through different mechanisms, such as the Facebook Timeline, Facebook Graph Search, and Google social website indexing. For example, the Facebook Timeline organizes a user's content according to a linear time line that allows easy browsing into past content. The increasing accessibility of historical information and the legislation regarding the right to be forgotten (EU Commission, 2012) give new urgency to certain thought-provoking questions: How would people cope with embarrassing information that was posted before they came into the workplace? What type of information could children see about their parents in the future? How should societies react to widespread and long-term information permanence? With OSNs becoming a vital part of people's social, political, and commercial activity, the question of data permanence is having a tremendous impact on people's lives. Certain scholars view digital data permanence as a potential danger to people and even to society (Mayer-Schönberger, 2009), whereas others view it as an opportunity to enhance sharing and transparency (Jarvis, 2011). Although the debate surrounding the impact of digital data permanence is lively, it is generally based on very

limited information that does not take an empirical view into the specific contexts in which information is shared and used for long periods.

To understand the effect of data permanence in OSNs, it is useful to contrast two approaches: one that views time as a construct in managing representation and another that views time as a construct in managing privacy. Although privacy and self-representation are strongly related (Tufekci, 2008), the first approach views information as a positive construct in self-representation, whereas the latter largely views information as a negative construct that has the potential of limiting a person's freedom. Hogan (2010) conceptualized the longitudinal aspects of social media as a set of exhibition spaces in which the social media platform acts as a curator, presenting the users through the accumulation of their digital traces. Zhao et al. (2013) expanded Hogan's exhibition metaphor and Goffman's (1959) dramaturgical theory in the context of data permanence in OSNs by analyzing the results of a qualitative study of Facebook users. They concluded that time mediates between different types of online activities, placing content in different spaces of public display and private archive. Content creation, profile management, and friending activities serve as current self-representation, or *performance*; later, as time passes, it becomes the user's own *exhibition*, and last, after a longer period, the content becomes a *personal* archival space. In these works, time is used as a positive construct in building and maintaining the self-presented image, with content generally becoming more personal over time.

Privacy scholars, however, mostly tend to view the effect of data permanence as a negative construct that complicates privacy management over time. Mayer-Schönberger (2009) described the dire implications of persistent digital memory by highlighting how the current society is experiencing a shift from forgetting by default to remembering by default. This revolutionary new situation has the potential to raise privacy concerns because it is more difficult to control, access, and comprehend large volumes of historical information. Forgetting also has a societal function, which is shown in its acceptance of people's ability to evolve and change by forgetting past events, such as bankruptcies and previous romantic relationships (Blanchette & Johnson, 2002). The popularity of transient messaging applications, which allow the exchange of self-destructing messages (such as Snapchat and Kik), highlights the problematic nature of full information permanency in social settings.

Novotny and Spiekermann (2014) surveyed users' preferences regarding data permanence on the web and found that approximately 80% of participants were in favor of enabling the Internet to forget, using deleting, obscuring, or hiding old content. The reasons for these preferences are related to increased privacy concerns for historical information, less control over personal information over time, and dissociation from the original context of the information. OSNs are fundamentally different from the general web; therefore, information aging and privacy can be fundamentally different. An OSN is an environment in which users produce most of the information, information is shared using privacy controls, and sharing is carried out to a network that changes over time. Similarly, permanence can play a different role in OSNs than in reminiscing applications, which are geared toward a single user or a small number of users (Cosley, Sosik, Schultz, Peesapati, & Lee, 2012; Crete-Nishihata et al., 2012;

Sellen & Whittaker, 2010). Both the web in general and reminiscing applications are representing a certain platform. Therefore, to properly understand data permanence in OSNs, we need to examine the particular context of privacy management in OSNs.

Because aged information in OSNs is readily available long after the first time it was published, it can be seen in different contexts than originally intended. Several studies have shown that OSN users are consistently underestimating the size of the audience that can see what they share (Bernstein, Bakshy, Burke, & Karrer, 2013) and the extent to which the information is accessed through time and space (boyd, 2009; Gross & Acquisti, 2005; Rosenblum, 2007), which possibly leads people to lower their restraints and to feel protected from future consequences (Rosenblum, 2007). Several theoretical frameworks may explain and predict the impact of time on privacy. Nissenbaum's contextual integrity theory relates privacy to the norms that surround an information flow in a specific context. A privacy violation occurs when "participants who consider themselves acting in one capacity in one context are treated as if they are acting in another capacity in a different context" (Nissenbaum, 2009, p. 225). As time passes, people may not understand the original context, therefore destabilizing the contextual integrity of information sharing. Furthermore, without understanding the original context, it is more difficult for the user to understand the implications that can be derived from sharing a piece of information (Grudin, 2001). A somewhat similar approach is presented by the adaptation of the research of Palen and Dourish (2003) to Altman's (1975) boundary regulation theory, in which time is understood as a complicating aspect of the regulation of the ongoing boundaries between the individual and others. Therefore, when following these theories, we expect to see growing privacy concerns for older information, which is outdated and out of context.

Research suggested that users rely heavily on privacy controls in OSNs and routinely manage their privacy settings in order to express their preferences (Lewis, Kaufman, & Christakis, 2008; Pew Internet and American Life Project, 2010; Stutzman, Gross, & Acquisti, 2013; Stutzman & Kramer-Duffield, 2010). Comprehending privacy preferences and management becomes more complicated as information ages in OSNs. For example, Wang et al. (2011) found that users' common regrets in using social media are due to unforeseen or ignored future consequences. In our previous study (Ayalon & Toch, 2013) we asked participants to provide feedback on gradually older Facebook posts. We found that users' willingness to share social media posts drops from the time since the post was first published at a rate of approximately 20% when comparing current posts and 2-year-old posts. However, although users perceive older posts as less relevant, their preferences are much more nuanced: They also want to keep some of their older posts for reminiscence purposes (Bauer et al., 2013). Bauer et al.'s results reflect an insignificant effect of time on participants' preferences about the total audience for the post, the study's sharing preferences variable, when participants provided feedback on current and year-old posts. These studies exemplify the high variation in the results, which are dependent on the study method and on the period of retrospective time frame. Our study isolates the impact of time from other perspectives. First, we base our results on a controlled experiment, in which the time since publishing the post is controlled in a random, between-subjects

study. Thus, controlling time has the benefit of controlling for participants' learning and framing biases (Kühberger, 1998), which can arise from comparing posts when iterating over a gradually older list of posts. Because the participants are independent and were placed in each condition randomly, we can confidentially quantify the extent of the effect of time.

Another gap in our understanding of OSN data permanence is based on the partial explanations that arise from looking either through the lens of privacy or through the lens of self-representation. To be able to explain data permanence, privacy and self-representation models should be fused together to provide the reasoning and the context behind the phenomena. To account for both the positive effect of time (how users and networks use time to define experience) and the negative effect of time (complicating privacy controls), we combine self-representation theories (Hogan, 2010; Zhao et al., 2013), privacy theories (Gross & Acquisti, 2005; Mayer-Schönberger, 2009), and online social dynamics theories (Chen, Geyer, Dugan, Muller, & Guy, 2009) in this study to provide a holistic view of longitudinal OSN usage.

We present a controlled study that empirically investigates the privacy preferences and manifested behavior of Facebook users with regard to historical information. We frame our research as examining *retrospective privacy*: how sharing preferences are affected by the time that has passed since the information was published and the possible reasons behind the effect. Toward this aim, we conducted a between-subjects experiment in which we presented the participants with their old Facebook posts from three epochs in the past: 0–1 years ago, 1–2 years ago, and 2+ years ago, as well as a control condition that included all of the available times. The different conditions allowed us to isolate the preferences according to time and provide feedback based on a single time frame rather than on a comparison between different times. Similar to previous studies (Ayalon & Toch, 2013; Bauer et al., 2013; Zhao et al., 2013), we find that there is a negative correlation between willingness to share a post and the time that has passed since it was first published. However, we also find the extent of this effect and then describe how it changes with the publication time. We analyze the factors that govern the ties between time and the user's preferences: relevancy, self-representation, social interaction that is related to the post, life changes, changes in the user's social network, and the age of the user. Our results reveal that older social network content cannot be classified to either personal or public spaces. Rather, content gradually becomes part of a "hybrid space" that combines selective public exhibition with personal archiving.

Our main contributions are threefold: (a) We experimentally measure the relationship between the publication time and the privacy preferences of the information; (b) we analyze how information relevancy, self-representation, and social interaction affect sharing preferences of older content; and (c) we explore how the results can be used to guide the design of longitudinal privacy management interaction models. Specifically, we look at proactive mechanisms, such as predicting which posts are possible candidates for long-term archiving. Finally, we discuss the theoretical framework that is necessary for understanding the relationship between time and privacy and the future challenges that are related to longitudinal privacy.

2. RETROSPECTIVE PRIVACY

Our model aims to explain the temporal aspects of information aging that are related to sharing preferences in Facebook, which is currently the most popular and widely used OSN. The model takes a retrospective approach by examining the relationship between information-sharing preferences and the time that has passed since the information was first published. After time has passed, the user begins to have a retrospective view of the post and may have different privacy, sharing, and representation preferences than those that he or she had when the post was first published. We are interested in understanding whether there are changes in privacy preferences regarding older posts; if so, what are the changes and what are the reasons for these changes?

Based on the relationship between context, time, and privacy (Grudin, 2001; Nissenbaum, 2009; Palen & Dourish, 2003), we explore the possible effect of time. Information contextual integrity is affected by time because older posts are more distant from the original context in which they were published than are newer posts, and therefore people will be less likely to share them. We combine the contextual integrity theories (Nissenbaum, 2009) and the self-representation theories when using OSNs (Hogan, 2010; Zhao et al., 2013) to relate performance and privacy management. This broad observation leads us to introduce the first hypothesis:

H1: Older information is less relevant to a person's current self-representation in an online social network.

Taking into account the context of OSNs and the possible effects of time led us to more hypotheses. H1 refers to contextual integrity in the eyes of the information's publisher.

As time passes, the social context changes: Old contacts, for whom a particular post was meaningful, drop away (Chen et al., 2009); new contacts who may misinterpret or not care about the information also lead to a gap between the current context and past information. These possible changes lead us to define our next hypothesis:

Hypothesis 2: The users' willingness to share is lower for older information as it is perceived as less relevant to others.

Similarly, in addition to social context changes over time, and possibly the cause for such changes (Chen et al., 2009), life events occur as well. Zhao et al. (2013) exemplified how meaningful events are used as a transiting milestone between the performance and the exhibition regions. Moreover, because young users experience life changes more frequently than older users, such as graduation, relationship changes, and so on, they may be more influenced by the impact of time. As a result, age may be a factor in disclosing information preferences. These observations lead us to the following two related hypotheses:

Hypothesis 3: The users' willingness to share is lower when the user experiences more major life changes.

Hypothesis 4: Younger users are more affected by the impact of time than older users.

Last, because we hypothesize that time will negatively affect users' willingness to share, we hypothesize that it will positively affect users' behavior:

Hypothesis 5: Users' inclination to alter Facebook posts (delete, change, or restrict access) is higher for older posts than new posts.

We expect that users will spend more effort on matching their preferences with reality. We based this observation on previous studies, which showed that users take steps to enhance their privacy in OSNs (Lewis et al., 2008; Pew, 2010; Stutzman et al., 2013; Stutzman & Kramer-Duffield, 2010).

3. METHOD

3.1. Experimental Method

To examine the hypotheses, we conducted a between-subjects experiment that compared the approaches toward sharing aged content, according to time epochs. We ran several experiments during November 2012, which was approximately 10 months after the introduction of Facebook's Timeline (McDonald, 2011). Participants were randomly assigned to one of four condition groups, with each representing posts from a different time epoch: (a) 0–1 years: between the current time and 1 year ago; (b) 1–2 years: between 1 year ago and 2 years ago; (c) 2+ years: more than 2 years ago; (d) 0–2+ years: a control condition that included random posts from any time. We added a control group to establish a benchmark for testing the different conditions, with the expectation that the differences would be weaker in comparison to the benchmark than when comparing any two other conditions. We chose Facebook because it is the largest OSN (71% of all of the U.S. adults online use Facebook), and it has been popular long enough for people to be active for at least 2 years on Facebook (Pew, 2013a). However, we did not want to examine specific blocks of time more than 2 years in the past in order to avoid a bias that may be caused by the many changes in the user interface that have occurred throughout the years.

For each condition, five posts that were published by the participant were randomly selected from the users' Facebook account and were presented to the user, one post at the time, to avoid a selection bias. A randomized post could be any of the following: status update, photo, video, geographical check-in, or link (but not comments or likes made on other posts). To implement the experiment, we built a custom web application that interacts with the Facebook Graph Application-Programming Interface to find and present past posts. The study was granted authorization by the Institutional Review Board.

FIGURE 1. The Users' Study Demographics and Essential Details, on a Per-post Basis.

Condition (epoch)	Number of Participants	Number of Posts	Mean time (in days) since publishing	Mean (SD) of willingness to share	Mean (SD) of relevance
0 – 1 years	82	380	171	4.12 (1.22)	3.55 (1.51)
1 – 2 years	55	255	545	3.71 (1.35)	3.35 (1.47)
2+ years	59	289	1072	3.51 (1.34)	2.95 (1.54)
0 – 2+ years	76	380	594	3.77 (1.36)	3.44 (1.52)
Total	272	1304	567		

The course of the experiment was as follows: Adult Facebook users were recruited by using Amazon's Mechanical Turk (MTurk), a crowdsourcing service that has become a popular recruiting method in human–computer interaction research (Kittur, Chi, & Suh, 2008). The MTurk task description did not directly mention privacy in order to avoid biasing our participant base by attracting people who are more sensitive to privacy concerns, which is similar to the technique in the study by Acquisti and Grossklags (2005). Participants who qualified for our experiment had to be in the United States, to control for language proficiency, and to have at least 95% positive MTurk feedback. The participants were asked to log in to their own Facebook account and allow the application to search for past posts. After these two steps, the participants were randomly assigned to one of the four conditions and asked to answer a questionnaire regarding a set of posts. The structure of the questionnaire included four parts: screening, demographics, sharing preferences, and general Facebook usage. Figure 1 displays the conditions' compositions in terms of time since the posts were published. To control for potential biases that result from an unequal number of participants between the conditions, we removed participants who would not have been valid for all conditions, due to not having enough old posts. The sharing preferences questionnaire is presented in Appendix A.

The sharing preferences questionnaire was presented in the context of a single post, which was randomly selected from the user's Facebook account, according to the time frame condition of the participant. The post included the original text, link, and photo or video of the post, as well as the Facebook likes and comments that the post received. The statements were rated on a 5-point Likert scale. The survey included two statements about intended behavior: "I am considering changing the status' content or deleting it in the future." and "I am considering hiding the status from some of my Facebook friends." Five other statements were geared toward evaluating both the privacy and self-representation hypotheses: (a) "I am satisfied with the status." (b) "The status is relevant today." (c) "The status may interest my Facebook friends today." (d) "If I published the status today, it will be a good representation of who I am today." and (e) "I would like the status to be seen in my timeline." The phrasing of the statements was based on Braunstein, Granka, and Staddon (2011), who evaluated suggestions for indirect privacy questions.

The questionnaire also included questions that elicited further descriptions of posts that would be used as control variables. The participants were asked to categorize the post's content, which was found to be a factor that affected sharing preferences (Wang et al., 2011; Zhao et al., 2013). Participants were able to choose only one option; the phrasing was "the status mostly includes": with the following possible answers, for example: general information about themselves, a post from a Facebook application, expression of personal feelings, information about other people. More possible answers can be found in Appendix A. Next, participants were asked to report their current social circle status and life changes that relate to the post. The social circle question was phrased as follows: "Review the people who commented or liked or shared the status. What best describes your relationship with them today?" The following are several possible answers: "I am in touch with few of them." and "I am in touch with all of them." The last questions in this section surveyed the existence of life changes. The participants were asked to rank their agreement with the following statements, which reflect the major personal and professional life changes that they experienced: (a) "Since publishing the post, I had major changes in my personal life (new relationship, new baby, moved to a new town or state, etc.)." and (b) "Since publishing the post, I had major changes in my professional life (switched to a new job, finished college, etc.)."

Along with collecting the participants' answers, the application automatically collected additional information about the participants and about the posts, including the number of likes, the number of comments, the date of publication, the post's type (status, link, video, photo, geographical check-in), and the privacy settings for the post (i.e., with whom the post was shared). The application collected information about the participant's number of Facebook friends. The post's content type, content category, its privacy settings, and the user's demographic information were later used as control variables. Except for the post's privacy settings, these variables were later treated as covariates. The post's privacy settings did not always exist; therefore, we excluded this variable from the model.

We recruited adult Facebook users through Amazon's MTurk. We screened for participants who were 18 or older and had been Facebook users for at least 2 years. We ran the experiments in 11 separate batches in order to facilitate the recruitment of an appropriate amount of participants. When combining these batches of experiments, the average payment for completing the survey was $0.65. There were slight variations in payment for different MTurk batches in an effort to find a reasonable compensation rate, ranging from $0.60 to $0.70 per survey. The survey took approximately 23 min to complete (*Mdn* = 8.35 min), and our compensation rate was approximately $1.50 an hour, which is higher than the median hourly reservation wage (Horton & Chilton, 2010; Paolacci, Chandler, & Ipeirotis, 2010). In regard to gender, 61% of the participants were female, 38% were male, and two participants did not reveal their gender (1%). Eighty-four participants were between the ages of 18 and 24 (31%), 119 participants were between the ages of 25 and 34 (43%), 60 participants were between the ages of 35 and 54 (22%), and nine participants were 55 or older (4%).

To determine whether our participants carried out the task according to the instructions, we used a double question screening task (Downs, Holbrook, Sheng, & Cranor, 2010) and paid attention to inconsistencies between similar questions. As a result, we removed the data of three participants. Due to differences in the number of participants between the conditions, we ensured unbiased selection by removing from all conditions those participants who reported having a Facebook account for approximately 2 years (22 participants). Thus, the remaining participants reported having their Facebook account 3 years or more. Sixty-three posts were removed as a result of a fault in identifying the time span; these posts were found in our data verification process carried out before the analysis. After checking the answers and filtering out empty and arbitrary answers, we removed 241 posts out of 1,545 total posts and the data of 37 participants out of 309 total participants.

3.2. Data Analysis

Our data included repeated measures by the same participant. Therefore, we employed a linear mixed effect (LME) analysis to take into account the random effects in our model (Winter, 2013). We tested all of the dependent variables for normal distributions, according to the Shapiro-Wilk test, and found that none of the variables were distributed normally ($p < .001$). However, the LME analysis has been shown to be robust and effective, even with non-normal, nonparametric variables (Howell, 1997, p. 321). We performed several LME analyses and used the posts' and participants' covariates. In the Results section, we compare the full model to the null model. The null model included the covariates, and the full model included the explored variables in addition to the covariates. The p values were obtained by likelihood ratio tests of the full model with the effects and were compared to the null model.

In addition to the LME analyses, we performed several other tests when analyzing the impact of time over willingness to share. To directly compare the groups in the different conditions, we first averaged each participant's score that was created by the repeated measures, which allows us to test the data on the basis of the participant and create independent observations. We used three nonparametric tests, which were based on the type of analysis: (a) The Spearman correlation test was used to test correlations (marked as ρ in the Results section); (b) the Kruskal-Wallis rank sum test was used to compare the results between groups (marked as H); and (c) the Siegel and Castellan (1988) test, which is a post hoc test that finds which groups are significantly different, was used after performing the Kruskal-Wallis test. We present the effect size for each significant comparison by using the following formula: $r = z/\sqrt{n}$, where 0.1–0.3 reflects a small effect size, 0.3–0.5 reflects a medium effect size, and greater than 0.5 reflects a large effect size. The data were analyzed using *R*, with the pgirmess package (Giraudoux, 2014) being used to perform a Siegel and Castellan post hoc test, and the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) being used for the LME analysis.

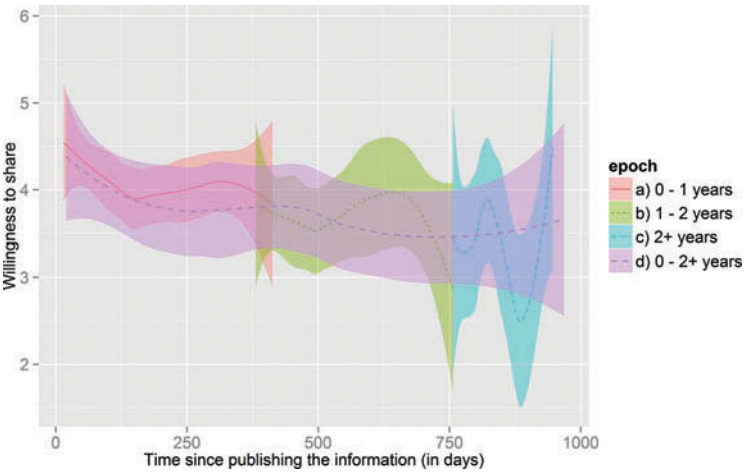
4. RESULTS

4.1. Impact of Time

We begin our analysis by examining the relationship between the participants' willingness to share their published posts and the time that has passed since the original publication date. For each post, the participants were asked to rank on a scale from 1 to 5 their agreement with the following statement: "I would like the status to be seen in my timeline." We regard their answer as their willingness to share. After averaging each participant's willingness score, the Kruskal-Wallis test was used and reveals that there is a significant difference in the willingness to share between the conditions (epochs; $H = 15.82$, 2 *df*, $p < .001$). A Siegel and Castellan post hoc test shows a significant difference between the following pair of conditions: 0–1 year and more than 2 years ($p < .05$, effect size $r = .315$). The median scores and interquartile range (IQR) for willingness to share, for all of the groups, from the 0–1 condition to the 2+ condition, are as follows, respectively: 4.2 (IQR = 1.4), 3.6 (IQR = 1.4), and 3.4 (IQR = 1.4). Over the 2 years, willingness to share decreases by approximately 20%, with a declining curve that reveals a decrease of 15% in the 1st year and a 5% decrease in the 2nd year. Figure 2 depicts the regression curves of willingness to share versus time, based on the entire data set (all of the epochs, including the control group). We created a regression curve that reflects the averaged willingness to share score per participant by using the loess-smoothing method. The colored region reflects a 0.95 confidence interval.

We perform an LME analysis of the relationship between willingness to share and the time that has passed since the post was published. We find a significant difference between our full model, which contains the effects of the epochs, and the null

FIGURE 2. Willingness to currently share facebook information versus the time that has passed since the information's first publication (in days).



Note. **The model is calculated based on the average for each participant.**

model, which does not contain the effects ($p = .021$). The overall marginal R^2 of the full model is 0.143. In the fixed effects level, we find that the epoch from which the posts were fetched significantly affected the willingness to share in both conditions: (a) 1–2 years compared to 0–1 year ($\beta = -0.303$, $p = .052$) and (b) 2+ years compared to 0–1 year ($\beta = -0.421$, $p = .008$), which shows that willingness to share decreases between the epochs. In regard to the postrelated covariates, participants who were in touch with all of those who had commented on or liked the post were more willing to continue to share it, when compared to those who were not in touch with any of those people ($\beta = 0.363$, $p = .05$). As detailed in Figure 3, other significant factors included the category of content, in which personal content (i.e., feelings and political views) were found to have a greater impact on willingness to share. The content categories are compared to a baseline category (posts published by Facebook's third-party applications on behalf of the user), which was included at the overall variability (intercept). The number of likes has a significant but weak positive effect on the willingness to share. Other covariates did not have a significant effect: the type of posts (i.e., link, photo, or video), the number of comments, and the participant's age and gender.

The decrease in willingness to share occurs in all conditions, as well as in the control group. To verify this result, we performed correlation tests to ensure similar results of the willingness to share between the three conditions when compared to the control group. The decrease in willingness to share over time, in terms of the correlation test, when using the three epochs' data is represented by $\rho = -0.302$ ($p < .0001$). When performing a correlation test by using the control group's data, the results show a decrease as well that is slightly more moderate: $\rho = -0.261$ ($p = .023$). This provides additional validation to the participant assignment process, providing additional strength to the claim that time since publishing the information is the separator between the conditions.

However, preferences do not translate into declared behavior, rejecting our fifth hypothesis. We asked the participants about their intended behavior in regard to deleting and hiding the post in the future. The phrasings of the statements were as follows: (a) "I am considering changing the status' content or deleting it in the future." and (b) "I am considering hiding the status from some of my Facebook friends." For both cases, we performed an LME analysis. However, we could not draw conclusions about these variables, as the full models were not significantly different from the null model. Therefore, we performed Kruskal-Wallis analyses, and for both cases we did not see significant differences between the conditions. A possible reason for this result could be the multiple roles of past posts, which sometimes serve as part of a "Personal Region" (Zhao et al., 2013) and sometimes as part of an "Exhibition Region." This dual role could explain users' unwillingness to commit to changing the post's content or privacy settings. The difficulties users face in controlling their OSN privacy (Bernstein et al., 2013; Madejski, Johnson, & Bellovin, 2011) can also explain why users are unlikely to make changes to their content sharing settings.

FIGURE 3. Linear Mixed Effect Model That Is Used for Predicting Willingness to Share, $R^2 = 0.143$.

	β	SE	t Value	Pr(>|t|)
(Intercept)	3.382	0.472	7.165	< **0.001**
Epoch: 1 – 2 years	−0.303	0.155	−1.958	**0.052**
Epoch: 2+ years	−0.421	0.158	−2.660	**0.008**
Changes occurred	−0.159	0.114	−1.395	0.163
Type: Check-in	−0.546	0.457	−1.194	0.233
Type: Link	0.132	0.339	0.389	0.698
Type: Photo	0.097	0.336	0.288	0.774
Type: Status	−0.295	0.345	−0.855	0.393
Who comment: All	0.363	0.185	1.966	**0.050**
Who comment: Few	−0.292	0.232	−1.259	0.208
Who comment: Most	0.180	0.195	0.925	0.355
Who comment: No one commented	−0.239	0.178	−1.338	0.181
Who comment: Some	0.063	0.224	0.282	0.778
Age: 25–34	0.023	0.146	0.158	0.875
Age: 35+	0.159	0.171	0.929	0.354
Gender: Male	−0.165	0.128	−1.292	0.198
Gender: Unspecified	−0.296	0.862	−0.344	0.732
No. of likes	0.035	0.013	2.692	**0.007**
No. of comments	−0.028	0.017	−1.637	0.102
Content: Feelings	0.833	0.264	3.151	**0.002**
Content: General self-information	0.833	0.261	3.190	**0.001**
Content: Other	0.608	0.259	2.351	**0.019**
Content: Other people	0.652	0.279	2.337	**0.020**
Content: Political view	0.794	0.365	2.178	**0.030**
Content: Product	0.544	0.300	1.813	0.070
Content: Question	−0.041	0.504	−0.082	0.935
Content: Relationship	0.749	0.302	2.484	**0.013**
Content: Request	−0.749	0.502	−1.491	0.136
Content: Work	1.136	0.360	3.158	**0.002**

4.2. Relevancy and Representation

To find nuanced explanations of longitudinal sharing, we analyze the participants' approaches toward different aspects of the historical post. We asked the participants about their satisfaction with the post and its relevancy. They were asked to rank their agreement with the following statements on a scale from 1 to 5: "I am satisfied with the status" (Satisfaction) and "The status is relevant today" (Relevancy). We performed LME analysis per each variable. For Satisfaction, the full model and the null model are not significantly differed. For Relevancy, the difference between the models is significant ($p = .009$, full model marginal $R^2 = .131$). The model shows a significant effect between the most recent epoch, 0–1 years, and the oldest one, 2+ years ($\beta = -0.56$, $p = .002$), showing that a post becomes less relevant with time. The entire model is represented in Appendix B.

Further analyses revealed that a post's irrelevancy has a major effect on sharing preferences. In a separate question, we asked about why a participant would consider hiding a post: "If you are considering hiding the status from your Timeline, what

would be the main reason?" The results pointed to irrelevancy as the major reason. In 284 posts of 1,304 total posts (22%), the participants stated that they would consider hiding the post. In 63% of those cases, participants chose irrelevancy as the main reason. A second reason was a change in the participants' point of view since they published the post (38 of 284, or 13%). Other reasons were as follows: the content seems inappropriate (9%), the content can offend the participant's friends (5%), and "other" (10%).

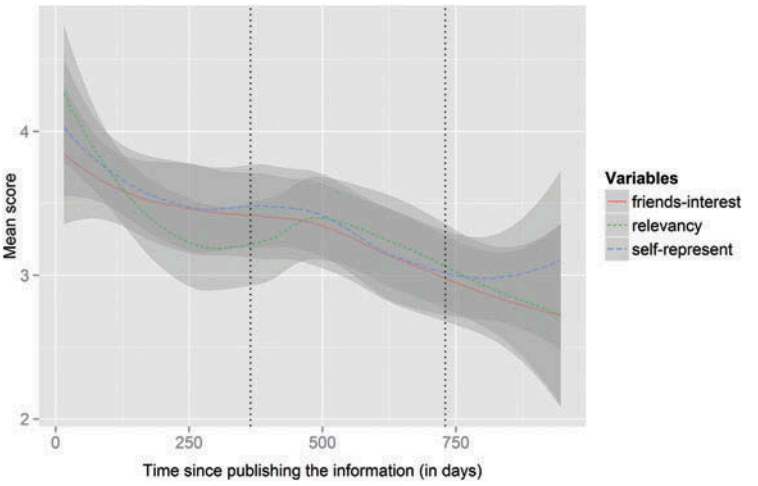
We compare the relevancy of the post to its audience and the relevancy of the post to the poster, meaning the one who published it. We explore whether one aspect of relevancy is affected differently by the impact of time. To test our first and second hypotheses, we test the following participants' perceptions regarding the following issues: (a) *Self-representation*: The participants were asked whether the post is a good representation of them ("If I were to publish the status today it would be a good representation of who I am today.") and (b) *Friends' interest*: We asked the participants whether they think that their Facebook friends would find the post interesting ("The status may interest my Facebook friends today.").

We use two LME analyses to study these aspects. In the self-representation analysis, we find a significant difference between our full model and the null model ($p = .013$, full model marginal $R^2 = .132$). The model, presented in Appendix C, shows a significant effect between the most recent epoch of 0–1 years and the oldest one, 2+ years ($\beta = -0.507, p = .003$), which shows that a post becomes less representative over time. The post's type significantly affected self-representation in one case, with check-ins being perceived as less representative. Self-representation was higher for posts that had more likes ($\beta = 0.056, p < .001$), and it was generally higher for older participants who were age 35 or older ($\beta = 0.501, p = .008$). The category of content was found to be significant as well, with content related to personal aspects significantly affecting self-representation. Content related to nonpersonal aspects, such as opinions on products, may point to a less personal post; therefore, their impact over either self-representation or willingness to share is weaker. Other fixed variables were not significant: the relationship between the post's publisher and those who commented on the post, the number of comments for the post, and the participant's gender.

The friends' interest LME results are similar to those for self-representation. Here, we also find a significant difference between our full model and the null model ($p = .003$, full model marginal $R^2 = .163$). In the fixed effects level, the model reveals a significant effect between the most recent epoch and the oldest one ($\beta = –0.590$, $p = .001$), which demonstrates that a post becomes less interesting to the participant's friends over time. Other fixed effect results are similar to the self-representation LME and are presented in Appendix D.

Figure 4 depicts the relationship between relevance, self-representation, and friends' interest. Generally, the variables are strongly correlated and decline over time. As a post ages, participants perceive it as being less representative of themselves, less relevant, and less interesting to their friends. We also find positive correlations between

FIGURE 4. The mean scores for friends' interest, the post's relevancy, and self-representation versus the time that has passed since the post was first published (in days).



willingness to share and all of these variables (relevance: $\rho = 0.6$, self-representation: $\rho = 0.672$, friends' interest: $\rho = 0.652$; $p < .0001$). Our results provide additional validation to our (Ayalon & Toch, 2013) and Bauer et al.'s (2013) previous results and, more importantly, offer nuances for the quantification of the rate of decline. Among these variables, the rate of decline is either the same or weaker for the 1st year and becomes larger during the 2nd year. For relevancy and friends' interest, the rate of decline for both the 1st and 2nd year is approximately 9% and 14%, respectively. For self-representation, the rates are 6% for the 1st year and 14% for the 2nd year. Relevancy, self-representation, and friends' interest decline in a steady or in an increasing rate, unlike willingness to share, for which the rate of decline weakens over time.

4.3. Social Circles

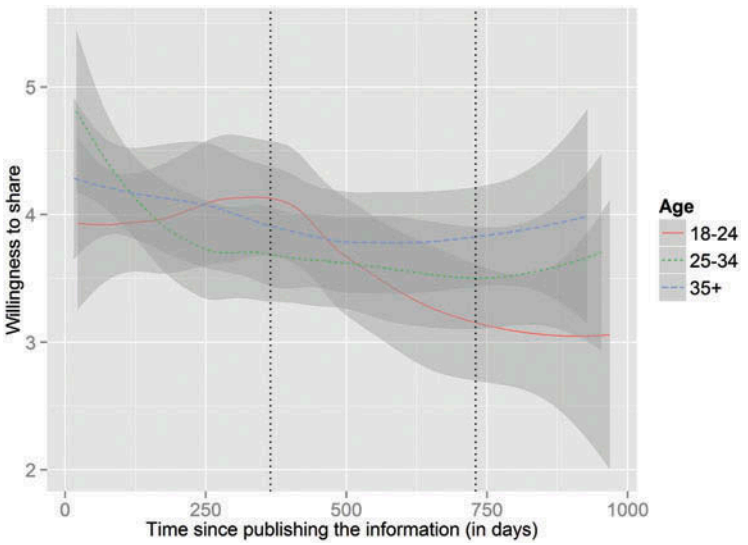
In the next set of results, we examine the relationship between social circles and impact on sharing perceptions. For each post, we asked the participants to examine those who had commented or liked the post and to choose the best description of the relationship between themselves and these people. The exact phrasing was "Please review the people who commented or liked or shared the status. What best describes your relationship with them today?" The answer was given on a scale that ranged from *I'm not in touch with them* to *I'm in touch with* ***all*** *of them* and included the following options: few, some, and most. The objective of this question is to explore the relationship between the audience of the post and the post's publisher. Moreover, as Gilbert and Karahalios (2009) showed, tie strength is correlated with OSN interaction. Therefore, analyzing the changes in the relationship with stronger relations might be more meaningful than with other Facebook friends. As presented in the first LME

analysis in Section 4.1, Figure 3, we found that only one relationship type was significant. When the participant is in touch with all of those who commented or liked the post, he or she is more willing to share a post than when he or she is not in touch with any of them ($\beta = 0.363, p = .05$).

4.4. Impact of Participants' Age

In our fourth hypothesis, we aim to understand the relationship between the user's age and the impact of time. Accordingly, the participants were divided into three age groups: 18–24, 25–34, and 35 years old or older. We united the two groups 35–54 and 55 years or older into one group in order to create groups that are similar in size (see Section 3.1). We hypothesized that younger users would be more affected by the impact of time than older users. We performed the LME analyses in which we fixed the age group per each analysis. However, because one model out of three was significant, we could not draw a conclusion that refers to the impact of time per each age group. Therefore, we performed a post-priori correlation test per each age group. The tests show a significant decrease in willingness to share over time for the first two age groups (18–24: $\rho = -0.297, p = .006$; 25–34: $\rho = -0.320, p < .001$). For the third age group (35 years or older), we did not find a significant correlation ($\rho = -0.172$, $p = .154$), leading us to conclude that older users are less affected by the impact of time. Figure 5 demonstrates the difference between the age groups by showing that the decrease in willingness to share is steeper for the younger participants.

FIGURE 5. Willingness to share, plotted by three age groups, versus the time that has passed since the post was first published (in days).



Note. **The model is calculated based on the participant's average willingness to share score.**

4.5. Life Changes

Last, we investigate our third hypothesis, in which we explore the impact of life changes on willingness to share. The participants were asked to rank from 1 to 5 their agreement about the occurrence of major life changes since the post was published with regard to two separate statements: (a) "Since publishing the post I had major changes in my personal life (new relationship, new baby, moved to a new town or state, etc.)" and (b) "Since publishing the post I had major changes in my professional life (switched to a new job, finished college, etc.)." We find a strong positive correlation between the number of life changes and time (personal life changes: $\rho = 0.464$, $p < .001$; professional life changes: $\rho = 0.398$, $p < .001$). We created two groups by combining the two statements and categorizing posts into a group in which *changes occurred* and a group in which *no changes occurred*. When controlling for the age group, life changes occurrence has a significant effect on willingness to share.

We performed three separate LME analyses, in which we fixed the age group per analysis, testing the interaction effect of life changes and age on willingness to share. Only in the case of the second age group (25–34) did we find a significant difference between the full model and the null model ($p = .034$). The overall marginal R^2 of the full model is .195. The model shows that when life changes occurred, the willingness to share decreases ($\beta = -0.370$, $p = .034$). Other significant covariates were number of likes and several content categories. In this analysis time was found insignificant.

To explore both the difference founded between the age groups and the relationship between age and life changes, we performed another post hoc analysis. We assumed that older users might have a smaller number of major changes that recently occurred, and therefore time may have a weaker effect on willingness to share for this group. For example, events that are related to changing social circles, such as graduation, have a higher chance of occurring in the more distant past for this group. Per each life changes type, personal and professional, we performed a Kruskal-Wallis test between the age groups, checking whether younger users experience more life changes. For both types the results show significant differences between the age groups, with more life changes occurring for younger users (professional: $H = 13.585$, 2 *df*, $p = .001$; personal: $H = 19.985$, 2 *df*, $p < .001$). For both life changes types, post hoc Siegel and Castellan (1988) tests show a significant difference between the youngest users (18–24) and the oldest users (35 or older) and between the medium-age users (25–34) and the oldest users (professional: 18–24 and 35+: $p < .05$, $r = .284$; 25–34 and 35+: $p < .05$, $r = .205$; personal: 18–24 and 35+: $p < .05$, $r = .356$; 25–34 and 35+: $p < .05$, $r = .215$). For professional life changes, the median life changes scores, from the youngest users (18–24) to the oldest (35 or older), are 3.0 (IQR = 2.688), 2.2 (IQR = 2.7), and 1.0 (IQR = 1.9). For personal life changes, the median life changes scores, from the youngest users to the oldest, are 3.2 (IQR = 2.2), 2.4 (IQR = 2.8), and 1.25 (IQR = 2). Also, the LME analysis that is described in Section 4.2 (Appendices C and D) shows that older participants perceive posts as more representative and as more interesting to friends than do younger users (representative: $\beta = 0.501$, $p = .008$; friends' interest: $\beta = 0.437$, $p = .02$).

5. DISCUSSION

The aim of this study was to evaluate retrospective privacy in OSNs in an experimental setting that included multiple years of retrospective analysis. In our study, we examined long-term sharing through the lens of privacy, self-representation, performance, and social dynamics. Our findings show that within the framework of retrospective assessment of old OSN content, time has a negative influence on the user's social circle's interest in the content and the content's relevancy (Hypothesis 1). In particular, the relevancy of historical information, or how well the user's social network relates to the content, positively influences the likelihood that a user would like to share it (Hypothesis 2). We show that life changes are correlated with a lower likelihood of sharing (Hypothesis 3) but with a relatively weak effect, which is significant only for participants from the second age group (25–34). Our findings suggest that younger users—who have more frequent changes in life—are more sensitive to temporal effects than older users and are less likely to share older posts (Hypothesis 4). However, we see that preferences are not easily translated into intentions, and although we saw that users report higher intentions to change or hide older information, the results were not significant (Hypothesis 5).

Our findings show that 93% of our participants were willing to share a recent post on Facebook (i.e., averaged willingness to share score was 3 or higher), 85% were willing to share a post that was published between 1 year and 2 years ago, and 75% were willing to share a post that was published 2 years ago or earlier. These results strengthen previous observations by Zhao et al. (2013), Ayalon and Toch (2013), and Bauer et al. (2013) by adding a between-subjects empirical analysis, in which time is controlled and the retrospection period is up to 3 years long. Our results provide additional support to these previous findings, but with a larger effect: Compared to our previous study (Ayalon & Toch, 2013), in which we found a correlation coefficient of $\rho = -0.19$ between time and willingness to share, in this study we find a coefficient of $\rho = -0.302$, with a comparable time span, statement phrasing, and analyses. We believe that the difference is mainly due to the random assignment to epochs and the between-subject design. More importantly, our findings show that the effect of time is not constant and linear. Rather, the effect of time on privacy approaches and willingness to share declines by 15% at the 1st year and by 5% at the 2nd year. It is the opposite case for self-representation, in which the decline rate increases at the 2nd year: 6% decline rate for the 1st year and 14% for the 2nd year. The validity of these results relies on both the random assignment to epochs and the random assignment of posts in each epoch.

5.1. Theoretical Discussion

Why have participants changed their sharing preferences? The analysis indicates two possible explanations: *staleness* and *anachronism*. In 42% of the surveyed posts from the 2+ years condition, we witness information staleness: A post with information that was 2 years old or older was defined as irrelevant by the participants, which significantly

impacted willingness to share the information. It is important to stress that the increasing irrelevancy is inherent, even if the post is not perceived as negative or embarrassing. These results correspond to the qualitative analysis of Facebook publishing behavior by Zhao et al. (2013), in which the passage of content from the social presentation space (exhibition) to the personal archive space is mediated by time. Our results quantitatively demonstrate that these spaces cannot be separable by time. Although there is a significant decline in willingness to share, in most of the cases, users still wish to share aged information, and in only 18% of cases of irrelevant posts did users wish to hide the post. Rather, time gradually generates a "hybrid space" that combines both public and personal regions. In this hybrid region, the preferences surrounding older posts can be traced to the content of the post, the current social context, and the difference with the social context at the time of the publication. The hybrid space is growing more heterogeneous with time, with increased variance in sharing preferences between posts as content grows older.

In a hybrid space, it is more difficult for users to maintain a shared context in which information can be communicated and understood. When social circles change, the information that was once shared with specific people in specific contexts is now an anachronism. For example, coworkers in a law firm might not understand the context in which party photos of drunk people were taken during college years. Our results reveal that 88% of the participants had reported at least one personal life change over the experiment's time span, and 36% of them had reported that this change had happened over the first 1-year period. In regard to social circles, we found that participants who were currently in relationships with all of those who commented or liked their post were more willing to share it. The social circles affect the norms of appropriateness that surround the information and may be perceived differently by contact that are not in one's social circle. The likelihood of sharing older posts declines because users cannot be sure that the information will be understood and viewed under the original appropriateness norms.

The relationship between the pace of life changes and sharing preferences is shown in the effect of age on sharing preferences. Older participants (35 years or older) were not significantly affected by time in terms of willingness to share, whereas younger participants were affected. We relate this observation to life changes and relevancy. As users grow older, they have fewer major changes in their lives (such as graduating, getting married, etc.). Therefore, the posts they are publishing are less likely to become irrelevant. We were not able to determine whether the effect of age is mediated or moderated by life changes occurrence. Further research is necessary to disambiguate the relationship between the two concepts. The usage habits of OSNs differ between the age groups, which have been previously observed (Pew, 2010).

5.2. Implications for Design

One of our motivations in this study was to inform the design of user interfaces and tools that help users better manage their privacy with long-term usage of OSNs. In the Results section, we present the gap between the sharing preferences of users

and their willingness to delete or change a post. Our results show that even in hybrid spaces, users do not tend to permanently change the content of historical information. Therefore, we conclude that solutions that completely and permanently change or delete content, such as the Web 2.0 Suicide Machine (Conley, 2010), or decaying mechanisms that delete the information after a certain amount of time (Barua, Kay, Kummerfeld, & Paris, 2011), would not appeal to the large majority of users as a solution for long-term retrospective privacy in OSNs. However, there might be subtler mechanisms for facilitating ongoing privacy management that do not delete the information. Our findings highlight how a "soft" retrospective privacy mechanism, such as *archiving*, can better serve users. In archiving, irrelevant information is moved to secondary storage that cannot be accessed directly by the user's current audience but can be accessed by the information's publisher. We suggest providing the user with an archiving option that fully restricts access to a post, unlike Facebook's hide feature, which removes content from the Timeline but does not remove content from other Facebook sources, such as photo albums.

Mayer-Schönberger (2009) argued for *data expiration dates* as a mechanism to reintroduce forgetting into the electronic domain. This mechanism triggers the destruction of information according to dates that are set up by users or regulators. Several research studies, based on cryptographic frameworks, may provide the technical foundations for such a technology, which makes it a viable possibility for retrospective privacy (Geambasu, Kohno, Levy, & Levy, 2009; Tang, Lee, Lui, & Perlman, 2010). Furthermore, several existing applications, such as Snapchat and Facebook's Slingshot, already implement an expiration date. However, the reluctance of our participants to delete old posts points to the conclusion that implementing an expiration date should take into account the particular characteristics of information aging in OSNs. We explain the difference in deletion preference as the difference between deletions in a performance region (such as in SnapChat) and deletions in a hybrid exhibition region (such as in Facebook posts). Although a performance region is inherently temporary in nature, a mixture of norms and uses characterizes a hybrid exhibition region. Deletion could be perceived as violating some of the uses and therefore was considered too harsh by participants.

Our results can be used to characterize the properties of expiration dates in OSNs while highlighting the challenges of implementing such a mechanism. For example, one of the basic questions regarding expiration dates is how to set the default expiration date. In our analysis, we can identify the tipping point with regard to sharing preferences. We performed the following regression analysis: Each 10 days we checked the percentage of participants who found the post irrelevant (irrelevancy score was 2 or below). The analysis result yielded the following equation: $y = 0.066x + 0.2$, in which y represents the proportions of negative participants' relevance scores and x represents the time since publishing the information, in years. The result suggests that 5 years is a good candidate for a default expiration date because beyond this point more than 50% of the posts were marked as irrelevant and not representative. The regression is based on the existing data, which mostly include posts that were published up to 3 years prior to the survey. However, although it is not based on posts from 5 years

ago, the regression helps us to predict a possible tipping point and can be rechecked in a similar experiment by using older posts to verify our results. Nevertheless, our results show that a one-size-fits-all solution, in which all of the messages are temporary for a preset length of time, might not be adequate for social networking. One important phenomenon to note is the variance in sharing preferences. A one-size-fits-all expiration date ignores the fact that 50% of posts were considered to be relevant by their owners after 5 years.

Allowing users to review aged content that they published can be a simple yet effective method for managing longitudinal privacy on OSNs. The diversity in sharing preferences hints that tools that allow users to review and quickly manage aged content might be useful. The Facebook Timeline is a good start for designing such a tool. Paradoxically, although the time line increases others' access to old information, it can also enhance awareness to the same information and control over that information. Our LME results highlight several attributes that may predict which posts the users may want to reconsider regarding their current privacy settings. We found that time since the post's publication, number of likes, current social circles, the occurrence of life changes since the post's publication, and the participant's age affected the user's willingness to share a post. Designing the actual actions that users can take regarding aged content raises several practical and theoretical questions. In 2015, Facebook added a new feature, "On this Day," which prompts the user to share posts from the past once again (Gheller, 2015). From our point of view, a mechanism like that could also be useful in archiving (rather than resharing) past information. Our results demonstrate that solutions that provide fine-grained control are preferable to solutions that permanently delete the information. However, as these solutions may add an additional burden on users (Benisch, Kelley, Sadeh, & Cranor, 2011), we argue for mechanisms that rely on some level of automation. One idea of such a mechanism is *contextual archiving*, which ensures that contacts would be able to see only the information of the user since they became friends with the user and that past information is inaccessible. The motivation behind such a limitation is to disable cases in which current friends might not understand the context of information that was published prior to their acquaintance with the user.

5.3. Limitations and Future Work

As we study long-term effects, the phenomena we present interact with other changes in OSN architecture and usage norms over time. In the last few years, Facebook has overhauled its interface (introducing Timeline), it has introduced new privacy controls (such as lists), it has changed the privacy defaults, and it has enrolled hundreds of millions of new users. The norms of publishing content and managing privacy have changed considerably as well (Pew, 2013b; Stutzman et al., 2013). Those changes make long-term privacy decisions even more complicated, as the context of publishing information is constantly changing. However, these interactions are largely controlled because we employed a retrospective method in which the time of the posts was experimentally controlled, and all conditions were evaluated simultaneously.

Other limitations to this study include the fact that our survey respondents were all recruited from MTurk, and thus our results may not necessarily be representative of the whole Facebook user population. In addition, because we recruited Facebook participants, our results may not be applicable to other social media, such as Twitter, or generally to social networks that do not have a Timeline feature (that provides fast access to old content). In addition, the questionnaire did not cover possible positive intentions to keep the posts, which may shed light on the way OSNs are managed as a shared archive. Last, our study and suggested tools did not cover third parties' ability to keep users' information once they publish it. Further studies are required to examine the fascinating impact of time on privacy and to unpack how hybrid spaces are perceived and used by people. Our study examines retrospective privacy that relies on backward retrospection of content that was published in the past. It will be interesting to explore future-facing privacy, where users will be asked about their time-related approaches and expectations when they actually post information to an OSN. Furthermore, future studies can investigate other factors that may impact users' willingness to share, such as the actual content of the post, the effect of user education interventions, nudging mechanisms (Wang et al., 2014), or peer behavior.

6. CONCLUSIONS

The public increasingly voices privacy concerns as more information is permanently stored in information systems and as access to this information becomes effortless. Many past research studies have dealt with privacy concerns and behaviors that mainly focus on users at the present time. However, current information technology allows information to be retained for long periods. This situation requires users to manage longitudinal privacy and to take into account the temporal effects of information aging on privacy. We believe that temporal privacy and topics that are related to the impact of time on the usage of information systems create one of the most crucial domains for HCI research. As the world shifts toward remembering digital information as its default, it is essential to understand users' concerns and to provide useful tools to mitigate these concerns.

Temporal effects on privacy are hotly debated and discussed by people in the media and in legislation forums around the world. However, there is almost no evidence to guide the discussion. To address this crucial gap, our study provides new insight into people's approaches and concerns and the potential impact of digital memory with social media. We analyze the results of a controlled study based on a technology that draws real information by real users on Facebook, the world's largest OSN ($n = 272$). In our study, we find that users are less willing to share old information in OSNs. We find several factors that further impact users' willingness to share, such as the post's relevance, the post's content, the post's audience, reactions to the post, changing social circles, and the occurrence of life changes.

Although data privacy concerns have constantly been monitored and investigated over the last four decades, it is becoming clear that digital memory will have a unique

impact on users' privacy behaviors. OSNs provide an experience of time in which two philosophical views of time coexist. OSNs form an interaction space in which "the past is never dead. It's not even past," in the words of William Faulkner (1951), but at the same time it is a space in which "the past is a foreign country: they do things differently there"—the opening sentence of *The Go-Between* by L. P. Hartley (1953). In OSNs, the past is easily accessible, but in many cases it provides an irrelevant and unrepresentative view of the user. To provide a positive user experience, OSN designs should take time into account and provide users with mechanisms for longitudinal information management and privacy.

NOTES

Background. This article is based on the M.Sc. thesis of the first author.

Acknowledgments. We thank Professor Noam Tractinsky, Professor Joachim Meyer, and Professor Michael Birnhack for their thoughtful comments on this article.

Support. This research was funded in part by a grant from the Israel Science Foundation grant no. 1116/12.

HCI Editorial Record. First received November 4, 2013. Revisions received September 10, 2014, and December 17, 2015. Accepted by Cliff Lampe. Final manuscript received June 7, 2016.

REFERENCES

Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. *Security & Privacy*, *3*(1), 26–33. doi:10.1109/MSP.2005.22

Altman, I. (1975). *The environment and social behavior*. Monterey, CA: Brooks/Cole.

Ayalon, O., & Toch, E. (2013). Retrospective privacy: Managing longitudinal privacy in online social networks. *Proceedings of the SOUPS 2013 Symposium on Usable Privacy and Security*. New York, NY: ACM.

Barua, D., Kay, J., Kummerfeld, B., & Paris, C. (2011). Theoretical foundations for user-controlled forgetting in scrutable long term user models. *Proceedings of the OzCHI 2011 Australian Computer–Human Interaction Conference*. New York, NY: ACM.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. Retrieved from http://CRAN.R-project.org/package=lme4

Bauer, L., Cranor, L. F., Komanduri, S., Mazurek, M. L., Reiter, M. K., Sleeper, M., & Ur, B. (2013). The post anachronism: The temporal dimension of Facebook privacy. *Proceedings of the WPES 2013 Workshop on Privacy in the Electronic Society*. New York, NY: ACM.

Benisch, M., Kelley, P. G., Sadeh, N., & Cranor, L. F. (2011). Capturing location-privacy preferences: Quantifying accuracy and user-burden tradeoffs. *Personal and Ubiquitous Computing*, *15*(7), 679–694. doi:10.1007/s00779-010-0346-0

Bernstein, M. S., Bakshy, E., Burke, M., & Karrer, B. (2013). Quantifying the invisible audience in social networks. *Proceedings of the CHI 2013 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Blanchette, J. F., & Johnson, D. G. (2002). Data retention and the panoptic society: The social benefits of forgetfulness. *The Information Society*, *18*(1), 33–45. doi:10.1080/01972240252818216

boyd, D. (2009). Why youth ♥ social network sites: The role of networked publics in teenage life. *Macarthur Foundation Series on Digital Learning–Youth, Identity, and Digital Media Volume*, 119–142.

Braunstein, A., Granka, L., & Staddon, J. (2011). Indirect content privacy surveys: Measuring privacy without asking about it. *Proceedings of the SOUPS 2011 Symposium on Usable Privacy and Security*. New York, NY: ACM.

Chen, J., Geyer, W., Dugan, C., Muller, M., & Guy, I. (2009). Make new friends, but keep the old: Recommending people on social networking sites. *Proceedings of the CHI 2009 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Conley, C. (2010). The right to delete. *Proceedings of the AAAI 2010 Spring Symposium Series*. Intelligent Information Privacy Management.

Cosley, D., Sosik, V. S., Schultz, J., Peesapati, S. T., & Lee, S. (2012). Experiences with designing tools for everyday reminiscing. *Human–Computer Interaction*, *27*(1–2), 175–198.

Crete-Nishihata, M., Baecker, R. M., Massimi, M., Ptak, D., Campigotto, R., Kaufman, L. D., & Black, S. E. (2012). Reconstructing the past: Personal memory technologies are not just personal and not just for memory. *Human–Computer Interaction*, *27*(1–2), 92–123.

Downs, J. S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010). Are your participants gaming the system?: Screening Mechanical Turk workers. *Proceedings of the CHI 2010 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

EU Commission. (2012). *Proposal for a regulation of the European Parliament and of the Council on the Protection of Individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*. Retrieved from http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf

Faulkner, W. (1951). *Requiem for a nun*. New York, NY: Vintage.

Geambasu, R., Kohno, T., Levy, A., & Levy, H. M. (2009). Vanish: Increasing data privacy with self-destructing data. *Proceedings of the USENIX 2009 Security Symposium*.

Gheller, J. (2015). *Facebook newsroom*. Retrieved from http://newsroom.fb.com/news/2015/03/introducing-on-this-day-a-new-way-to-look-back-at-photos-and-memories-on-facebook/

Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. *Proceedings of the CHI 2009 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Giraudoux, P. (2014). *pgirmess: Data analysis in ecology*. Retrieved from http://CRAN.R-project.org/package=pgirmess

Goffman, E. (1959). *The presentation of self in everyday life*. New York, NY: Anchor Books.

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. *Proceedings of the WPES 2005 Workshop on Privacy in the Electronic Society*. New York, NY: ACM.

Grudin, J. (2001). Desituating action: Digital representation of context. *Human Computer Interaction*, *16*(2), 269–286. doi:10.1207/S15327051HCI16234_10

Hartley, L. P. (1953). *The go-between*. London, UK: Penguin.

Hogan, B. (2010). The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society*, *30*(6), 377–386.

Horton, J. J., & Chilton, L. B. (2010). The labor economics of paid crowdsourcing. *Proceedings of the EC 2010 Conference on Electronic Commerce*. New York, NY: ACM.

Howell, D. C. (1997). *Statistical methods for psychology*. Belmont, CA: Wadsworth.

Jarvis, J. (2011). *Public parts: How sharing in the digital age improves the way we work and live*. New York, NY: Simon and Schuster.

Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. *Proceedings of the CHI 2008 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *75*(1), 23–55. doi:10.1006/obhd.1998.2781

Lewis, K., Kaufman, J., & Christakis, N. (2008). The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication*, *14*(1), 79–100. doi:10.1111/jcmc.2008.14.issue-1

Madejski, M., Johnson, M. L., & Bellovin, S. M. (2011). *The failure of online social network privacy settings* (Tech. Rep. CUCS-010-11). New York, NY: Columbia University.

Mayer-Schönberger, V. (2009). *Delete: The virtue of forgetting in the digital age (New in Paper)*. Princeton, NJ: Princeton University Press.

McDonald, P. (2011). *Timeline: Now available worldwide* [Blog post]. Retrieved from https://www.facebook.com/notes/facebook/timeline-now-available-worldwide/10150408488962131

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford University Press.

Novotny, A., & Spiekermann, S. (2014). Oblivion on the Web: An inquiry of user needs and technologies. *Proceedings of the ECIS 2014 European Conference on Information Systems*.

Palen, L., & Dourish, P. (2003). Unpacking privacy for a networked world. *Proceedings of the CHI 2003 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, *5*(5), 411–419.

Pew Internet and American Life Project. (2010). *Reputation management and social media*. Retrieved from http://www.pewinternet.org

Pew Internet and American Life Project. (2013a). *Social Media Update*. Retrieved from http://www.pewinternet.org

Pew Internet and American Life Project. (2013b). *Teens, Social Media, and Privacy*. Retrieved from http://www.pewinternet.org

Rosenblum, D. (2007). What anyone can know: The privacy risks of social networking sites. *IEEE Security & Privacy Magazine*, *5*(3), 40–49. doi:10.1109/MSP.2007.75

Sellen, A. J., & Whittaker, S. (2010). Beyond total capture: A constructive critique of lifelogging. *Communications of the ACM*, *53*(5), 70–77. doi:10.1145/1735223

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences*. New York, NY: McGraw-Hill.

Stutzman, F., Gross, R., & Acquisti, A. (2013). Silent listeners: The evolution of privacy and disclosure on Facebook. *Journal of Privacy and Confidentiality*, *4*(2), 7–41.

Stutzman, F., & Kramer-Duffield, J. (2010). Friends only: Examining a privacy-enhancing behavior in Facebook. *Proceedings of the CHI 2010 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Tang, Y., Lee, P. P., Lui, J. C., & Perlman, R. (2010). *Fade: Secure overlay cloud storage with file assured deletion*. *Proceedings of the ICST 2010 Conference on Security and Privacy in Communication Networks*. Berlin, Germany: Springer.

Tufekci, Z. (2008). Grooming, gossip, Facebook and MySpace: What can we learn about these sites from those who won't assimilate? *Information, Communication & Society*, *11*(4), 544–564. doi:10.1080/13691180801999050

Wang, Y., Leon, P. G., Acquisti, A., Cranor, L. F., Forget, A., & Sadeh, N. (2014). A field trial of privacy nudges for Facebook. *Proceedings of the CHI 2014 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., & Cranor, L. F. (2011). I regretted the minute I pressed share: A qualitative study of regrets on Facebook. *Proceedings of the SOUPS 2011 Symposium on Usable Privacy and Security*. New York, NY: ACM.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv*, 1308.5499.

Zhao, X., Salehi, N., Naranjit, S., Alwaalan, S., Voida, S., & Cosley, D. (2013). The many faces of Facebook: Experiencing social media as performance, exhibition, and personal archive. *Proceedings of the CHI 2013 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

APPENDIX A

The sharing preferences questionnaire
[POST APPEARS]

The status mostly includes:
() General information about myself (photo, location, etc.)
() Expression of my personal feelings
() Information about other people
() Information about my work or company
() Information about my relationships (family, romantic, etc.)
() Political views
() My opinion about a certain product/service
() A request for some kind of help (finding an apartment, looking for job, etc.)
() Post from a Facebook application
() A question
() Other, please specify

On a scale from 1 to 5, how much do you agree or disagree with the following statements? (1 – I highly disagree, 5 – I highly agree)
I am satisfied with the status ()
I am considering changing the status' content or to delete it in the future. ()
I am considering hiding the status from some of my Facebook friends. ()
The status is relevant today ()
The status may interest my Facebook friends today ()
If I would publish the status today it will be a good representation of who I am today ()
I would like the status to be seen in my time line ()

If you are considering hiding the status from your time line, what would be the main reason?

() It can offend my friends.
() It seems irrelevant.
() There is a change in my point of view since I had posted the status.
() It seems inappropriate that some of my Facebook friends will see it.
() I would not consider to hide the status from my time line.
() Other, please specify

Please review the people who commented or liked or shared the status. What best describes your relationship with them today?

() I am in touch with all of them
() I am in touch with most of them
() I am in touch with some of them
() I am in touch with few of them
() I am not in touch with them
() No one commented nor liked nor shared the status

Please check the post's publication date. On a scale from 1 to 5, how much do you agree or disagree with the following statements? (1 – I highly disagree, 5 – I highly agree)

Since publishing the post I had major changes in my personal life (new relationship, new baby, moved to a new town or state, etc.) ()
Since publishing the post I had major changes in my professional life (switched to a new work, finished college, etc.) ()

APPENDIX B

TABLE B1. Linear Mixed Effect Model That Is Used for Predicting a Post's Relevancy

	Estimate	SE	*t* Value	Pr(>|t|)
(Intercept)	2.947	0.551	5.343	< **0.001**
Epoch: 1–2 years	−0.164	0.181	−0.910	0.364
Epoch: 2+ years	−0.560	0.181	−3.091	**0.002**
Type: Check-in	−1.434	0.541	−2.650	**0.008**
Type: Link	−0.114	0.401	−0.283	0.777
Type: Photo	−0.257	0.398	−0.646	0.519
Type: Status	−0.959	0.408	−2.349	**0.019**
Who comment: All	0.377	0.219	1.723	0.085
Who comment: Few	−0.245	0.274	−0.895	0.371
Who comment: Most	0.096	0.230	0.417	0.677
Who comment: No one commented	0.083	0.211	0.392	0.695
Who comment: Some	−0.066	0.265	−0.249	0.803
Age: 25–34	0.059	0.174	0.337	0.737
Age: 35+	0.410	0.199	2.060	**0.041**
Gender: Male	−0.104	0.151	−0.689	0.491
Gender: Unspecified	−0.063	1.022	−0.062	0.951
No. of likes	0.053	0.016	3.421	**0.001**
No. of comments	0.014	0.020	0.690	0.490
Content: Feelings	0.925	0.313	2.956	**0.003**
Content: General self—information	0.682	0.309	2.206	**0.028**
Content: Other	0.426	0.306	1.389	0.165
Content: Other people	0.518	0.331	1.566	0.118
Content: Political view	1.238	0.431	2.870	**0.004**
Content: Product	0.628	0.356	1.767	0.078
Content: Question	0.384	0.597	0.643	0.521
Content: Relationship	1.010	0.357	2.830	**0.005**
Content: Request	−0.770	0.595	−1.295	0.196
Content: Work	1.032	0.426	2.424	0.016

APPENDIX C

TABLE C1. Linear Mixed Effect Model That Is Used for Predicting a Post's Self-Representation.

	Estimate	SE	t Value	Pr(>|t|)
(Intercept)	2.712	0.531	5.110	< **0.001**
Epoch: 1–2 years	−0.226	0.170	−1.324	0.187
Epoch: 2+ years	−0.507	0.171	−2.965	**0.003**
Type: Check-in	−1.274	0.522	−2.440	**0.015**
Type: Link	0.035	0.387	0.089	0.929
Type: Photo	0.049	0.384	0.127	0.899
Type: Status	−0.419	0.394	−1.064	0.287
Who comment: All	0.291	0.212	1.373	0.170
Who comment: Few	−0.238	0.265	−0.895	0.371
Who comment: Most	0.082	0.223	0.370	0.712
Who comment: No one commented	−0.121	0.204	−0.594	0.553
Who comment: Some	−0.031	0.256	−0.122	0.903
Age: 25–34	0.253	0.164	1.547	0.124
Age: 35+	0.501	0.188	2.670	**0.008**
Gender: Male	−0.179	0.142	−1.261	0.209
Gender: Unspecified	−0.164	0.963	−0.170	0.865
No. of likes	0.056	0.015	3.681	< **0.001**
No. of comments	−0.014	0.019	−0.722	0.470
Content: Feelings	0.916	0.301	3.047	**0.002**
Content: General self—information	0.800	0.297	2.694	**0.007**
Content: Other	0.520	0.294	1.765	0.078
Content: Other people	0.417	0.318	1.311	0.190
Content: Political view	1.043	0.416	2.509	**0.012**
Content: Product	0.589	0.342	1.721	0.086
Content: Question	0.630	0.577	1.091	0.276
Content: Relationship	0.974	0.343	2.837	**0.005**
Content: Request	−0.615	0.574	−1.071	0.284
Content: Work	0.849	0.410	2.070	0.039

APPENDIX D

TABLE D1. Linear Mixed Effect Model That Is Used for Predicting a Post's Interest.

	Estimate	SE	t Value	Pr(>|t|)
(Intercept)	3.092	0.524	5.905	< **0.001**
Epoch: 1–2 years	−0.252	0.169	−1.492	0.137
Epoch: 2+ years	−0.590	0.169	−3.484	**0.001**
Type: Check-in	−1.505	0.515	−2.922	**0.004**
Type: Link	−0.108	0.382	−0.284	0.776
Type: Photo	−0.261	0.379	−0.690	0.490
Type: Status	−1.023	0.388	−2.634	**0.009**
Who comment: All	0.057	0.209	0.272	0.785
Who comment: Few	0.022	0.262	0.085	0.932
Who comment: Most	0.098	0.219	0.445	0.657
Who comment: No one commented	−0.235	0.201	−1.169	0.243
Who comment: Some	−0.110	0.253	−0.433	0.665
Age: 25–34	0.083	0.162	0.510	0.610
Age: 35+	0.437	0.186	2.351	**0.020**
Gender: Male	−0.199	0.141	−1.410	0.160
Gender: Unspecified	−0.067	0.955	−0.070	0.944
No. of likes	0.060	0.015	4.047	< **0.001**
No. of comments	−0.009	0.019	−0.450	0.653
Content: Feelings	0.861	0.297	2.903	**0.004**
Content: General self—information	0.852	0.293	2.907	**0.004**
Content: Other	0.743	0.291	2.557	**0.011**
Content: Other people	0.925	0.314	2.948	**0.003**
Content: Political view	0.876	0.410	2.136	**0.033**
Content: Product	0.587	0.338	1.740	0.082
Content: Question	0.581	0.569	1.021	0.307
Content: Relationship	1.299	0.339	3.832	< **0.001**
Content: Request	−1.280	0.566	−2.262	**0.024**
Content: Work	0.880	0.405	2.174	0.030