"I just want to feel safe": A Diary Study of Safety Perceptions on Social Media

Elissa M. Redmiles, Jessica Bodford, and Lindsay Blackwell

Abstract

Social media can increase social capital, provide entertainment, and enable meaningful discourse. However, threats to safety experienced on social media platforms can inhibt users' ability to gain these benefits. Threats to safety – whether real or perceived – detract from the pleasure people get out of their online interactions and damage the quality of online social spaces. While prior work has individually explored specific threats to safety – privacy, security, harassment – in this work we more broadly capture and characterize the full breadth of day-to-day experiences that influence users' overall perceptions of safety on social media. We explore these perceptions through a three-week diary study (n=39). We contribute a novel, multidimensional taxonomy of how social media users define 'safety', centered around security, privacy, and community. We conclude with a discussion of how safety perceptions can be used as a metric for social media quality, and detail the potential for enhancing safety perception through community-enhancing affordances and algorithmic transparency.

Introduction & Background

People reportedly spend an average of 2 hours and 15 minutes a day on social media sites connecting with friends, sharing life updates, or playing games (AdWeek). This time can bring social media users great benefit: e.g., increasing social capital, enabling stimulating discourse, and assisting with information acquisition (Ellison, Steinfield, and Lampe 2007; Burke, Kraut, and Marlow 2011; Morris, Teevan, and Panovich 2010). However, experiences of threatened safety on social media can inhibit these gains (Gl´uer and Lohaus ; Kross and et al. 2013; Kim, LaRose, and Peng 2009; Holmes and O'loughlin 2014). Feeling unsafe is not only detrimental to users' well-being, but can also limit users' sense of self-efficacy – their belief that they can use technological controls to enhance their own safety – thus inhibiting their ability to defend against threats and improve their own social media experiences (Bodford 2017; Lee, Larose, and Rifon 2008).

Experiences of threat in specific categories – security, privacy, harassment – have been explored independently from each other in the context of social media (Madden and et al 2013; Holmes and O'loughlin 2014; Ellison et al. 2011; Litt 2013; Zhang, De Choudhury, and Grudin 2014; Wisniewski, Knijnenburg, and Lipford 2014), online dating (Stenson, Balcells, and Chen 2015; Gibbs, Ellison, and Lai 2011), email, and banking (Udo 2001; Furnell, Bryant, and Phippen 2007; Camp 2009). Studies of social media sites describe a number of threats identified by users, particularly around privacy settings, controls, and identity management (Bauer and et al. 2013). Studies of teens and social media emphasize offline safety in addition to privacy, especially situations of harassment and cyberbullying (Boyd 2014; Ashktorab 2018; Ashktorab and Vitak 2016).

Each of these studies raises a different element of threat, but the broader concept of online safety–that is, the full spectrum of both day-to-day and infrequent (but impactful) experiences that contribute to a user's overall sense of safety–remains unclear. In this work, we take a first step toward investigating the day-to-day experiences that affect social media users' perceptions of safety and threat. To do so, we conducted a microlongitudinal diary study of 39 U.S. users of Facebook, the most-used social media site in the U.S. (Smith and Anderson 2018) We asked participants to report safety experiences–specifically, salient moments when they felt either safe or threatened on Facebook–over a two-week period. To place these experiences in the context of participants' broader perceptions of online safety, we also asked participants to describe an experience of safety and an experience of threat that they recalled from any time in the past on any platform.

We specifically used the term 'safety' to allow users to define for themselves the concepts and affordances that contribute to their sense of feeling 'safe' (or unsafe) on a platform, rather than imposing a specific or more narrow definition of safety (such as 'security' or 'privacy'). While researchers rigorously define and distinguish between concepts like privacy, security, and harassment, everyday users may not. Indeed, in some non-English languages, 'safety' is a single word which encompasses all of these concepts. For example, although companies may promise to protect users' information from access by third parties, malicious players (e.g., hackers; disgruntled employees) may obtain that information anyway – which does not imply a breach of privacy, but rather a breach of security. By using the term *safety* we include both such violations, in addition to other potential perceptions of threats, such as those that relate to physical

safety (which may be especially relevant for social media users, given the relationship between social media and both online and offline harassment (Duggan et al. 2017)).

Finally, using the results of our diary study, we contribute a novel taxonomy of safety on Facebook. We find that the experiences that made participants feel safe or threatened fell into three broad categories: privacy, security, and community, with participants reporting a near-equal portion of experiences of each.

Our results suggest that individual experiences and perceptions of safety on social media sites are multifaceted, involving not only digital privacy, security, and harassment, but also offline safety and well-being and the upholding of community values. As such, studying, for example, social media harassment in isolation–without considering a user's overall experience with and perception of safety–may miss crucial context for understanding user behavior and improving platform design. We conclude by offering a preliminary framework for conceptualizing safety, discuss the potential for safety to serve as a multifaceted metric for assessing social media quality or health, and offer suggestions for algorithmic and interface design changes that can increase social media safety.

Methodology

We conducted our diary study, which was approved through our organization's research and ethics review process, from August to September 2017.

The primary aim of our research was to understand moments when people feel particularly safe or threatened. Because we hoped to understand day-to-day perceptions of safety we chose to use a diary study method. Our study uses a feedback diary study approach, in which participants captured each moment on their own time, without prompting on our part (Carter and Mankoff 2005).

Diary study methods offer advantages for examining day-to-day experiences such as those we wish to examine here. Diary studies allow a more naturalistic approach to data gathering than standard lab studies because participants can report experiences in their own contexts and on their own time schedule (Palen and Salzman 2002). Further, in a one-time laboratory or online survey setting, participants are forced to think of examples that stood out in their minds as especially uncommon, and may forget to mention more routine events that effect perception over time, but are not individually memorable. Furthermore, asking about experiences based on participants' memories introduces the confound of time, in which experiences with (for example) Facebook's site-wide changes to privacy settings in 2009 would now be obsolete in understanding broader security experiences on social media in the present year. We do collect such recollections *in addition* to day-to-day experiences in order to better understand the differences between day-to-day experiences of safety and long term experiences or perceptions – a benefit of the diary study method is that it affords us the unique opportunity to collect both recollections and experiences.

Recruitment We recruited participants to complete a screening survey via dScout (dSc). dScout is a commonly used tool for conducting diary studies via mobile phone (Winnick 2012). dScout maintains a panel of potential participants, who dScout recruits through a variety of methods (e.g., online advertisements, advertisements through frequent flyer programs, mailers, etc.) designed to ensure panel diversity – similar to non-probabilistic survey panels maintained by companies such as Qualtrics (Redmiles et al. 2017).

Our screening survey asked respondents how frequently they used each of the following seven applications or application types: Facebook, Instagram, Twitter, Snapchat, Personal Email, Online Banking, Dating Applications (order of options was randomized); on which of those applications they felt the safest; and, through an open ended question, why they felt safest using that application. Finally, respondents were asked a series of standard demographic questions, collecting their gender, race, age, income, and level of educational attainment. The survey was reviewed by survey experts to ensure that best practices were followed for obtaining valid, high-quality data (Redmiles et al. 2017).

From the 525 respondents who completed this screening survey, we selected 39 participants for our diary study. We selected participants who had (1) provided thorough answers to our survey questions – since these questions are similar to our diary study prompts, (2) reported using Facebook multiple times a week, such that they would be on the platform enough in a two-week period to report experiences without an artificial increase in platform use, and (3) were demographically diverse in terms of their age, gender, income, education, and race in an attempt to observe a variety of experiences [1]. Participants were paid $100 for their participation in the full study.

Protocol Our three-week diary study consisted of two parts: in the first we obtained general examples of participants' online experiences of safety and threat; while in the second we collected participants' day-to-day safe (and unsafe) experiences on Facebook.

In the first week, we asked participants to report a safe and an unsafe online experience that they recall had at any time in the past. Participants completed this first diary entry in the first week of data collection. We use this data to put participants' day-to-day experiences in context, and to explore the difference between day-to-day experiences and more memorable events. This first diary entry also served to acclimate participants to the focus of the study and to prepare them for the six remaining diary entries.

In weeks 2 and 3, participants were asked to provide three data points each regarding an experience that made them feel particularly safe or unsafe while on Facebook during the two-week study, for a suggested total of six data points per participant. After encountering an experience, participants navigated to the dScout platform and selected whether they wished to report an unsafe or safe experience. Participants saw the following prompt for recording their 60 minute ex-

[1] Prior work indicates that social media behavior and experiences may vary with demographics, and thus we sought demographic diversity to ensure that we covered a maximum set of possible experiences (Madden and et al 2013; Hargittai and Litt 2011; Redmiles 2018a).

perience description: “Please tell us what happened to make you feel particularly [safe/unsafe]. How did the experience start? Did it happen to you or someone you know? Was anyone else involved? What ended up happening?” For unsafe experiences, participants were asked to rate the intensity of the experience on a 10-point Likert scale: “Please use the scale below to indicate how INTENSE this experience was for you, where 0 is ”Not at all intense” and 10 is ”Extremely intense”.

Focus on a Single Platform: Facebook. In order to provide a deep exploration of day-to-day experiences of safety and threat, we focus primarily on a single platform. By considering one exemplar social media platform we allow for the collection of experiences around consistent, comparable affordances. We chose to focus on Facebook as our exemplar platform for three reasons: (1) it is the most highly used social media platform in the U.S. today (Center 2017), thus offering more opportunity for people to record relevant experiences within the study timeframe; (2) it has been relatively understudied, as compared to e.g., Twitter, especially with regard to security topics (Redmiles, Chachra, and Waismeyer 2018); (3) it has a high number of affordances similar to those of other platforms (e.g., stories like Snapchat and Instagram, Like and Resharing functionality similar to Twitter) thus allowing participants to discuss any number of components that may be relevant to research on other platforms. In our reporting of the results, we place these Facebook-specific findings in the broader context of the safety and threat experiences that participants described in week 1 – which occured on a diversity of social media and other platforms (email, banking, etc.).

Analysis To report experiences, participants recorded 60-second videos and answered a closed-answer survey question, when applicable (about experience intensity); the videos were transcribed by dScout. Data from the introductory, week one, diary study task was analyzed separately and with a separate codebook (which contained a subset of the codes contained in the Facebook-specific diary study codebook) than the two-week, Facebook-specific portion of the diary study. As there were only 78 transcripts to code for the first week portion of the study, all of the data was double coded; with the two researchers achieving a mean Krippendorff’s alpha of 0.89 and a percentage agreement of 0.97. For the Facebook-specific portion of the study, as there were 222 video transcripts, two researchers coded 20% of the 222 transcripts. They achieved a Krippendorff’s alpha (Hayes and Krippendorff 2007) of 0.96 and percentage agreement of 0.99. As their agreement was high and double-coding the remaining data would have taken significant quantities of time, one of the researchers then coded the remaining 80% of the data, following standard qualitative coding practices (Braun and Clarke 2006; Joyce 2013; McNally and et al 2017; Tausczik, Wang, and Choi 2017; Druin and et al 2010).

In the results, we report both numbers of experiences, or participants, and percentages of the whole, to avoid overstating generalizability. Participant quotes have been lightly edited for readability.

Finally, despite the small sample size, a post-hoc power analysis revealed that the mean intensity differences across themes were large enough for us to have sufficient power to analyze and compare the Likert scale intensity data across experiences using a one-way randomized ANOVA. For significant ANOVA results, we also report Cohen’s d to quantify effect sizes.

Participants

The majority of our potential participant pool (the 525 respondents to the screening survey) were female (70%), with a mean age of 34 (SD=10); we anticipate that this skew occurred because there are more women in the dScout panel generally, and additionally, research has found that women are more likely to use social media than men, perhaps influencing their interest in a study about social media (Center 2017). Our survey respondents were also more educated than the general U.S. population: 72% had at least a college degree compared with 30% of the U.S. population; this educational skew is also seen in the social media-using population in general (Center 2017). Our respondents were slightly wealthier than the general population, with 29% reporting a household income greater than $100K. They are nearly racially representative of the U.S., although fewer identified as Hispanic/Latino (10%) and more identified as Asian (11%) compared with 18% and 6% in the US, respectively. The modal mobile operating system in our sample was iOS, with 65% of respondents using an iPhone to complete the questionnaire.

From this potential participant pool, we selected 39 participants for the diary study, with the goal of diversifying our participant pool[1] and ensuring high quality responses. 19 diary study participants were male and 20 were female; 12 had a college degree or higher educational attainment; 9 had an income over $100K, 15 had an income between $35K and $100K, and the remainder earned less than $35K; and their racial identity approximately matched the distribution of the U.S. population. Eleven of these participants used Android phones and 28 used iPhones.

Limitations

The primary limitation of diary studies is the potential for demand effects – that is, for participants to over-observe due to their participation in the study (Scollon, Prieto, and Diener 2009; Iida et al. 2012; Palen and Salzman 2002). To mitigate these effects, we offered participants flexibility in reporting to help minimize demand characteristics: they were still compensated if they reported more or less than six experiences, or four safe and two unsafe experiences (participants reported an average of 5.8 experiences and 2.8 safe experiences reported). Diary studies offer an alternative set of benefits and limitations to interview studies, which have been more extensively used in prior research on social media privacy and harassment (Blackwell et al. 2018; Page, Kobsa, and Knijnenburg 2012; Strater and Lipford 2008). Interview studies suffer from more recall biases than diary studies, while diary studies suffer from stronger demand effects (Scollon, Prieto, and Diener 2009; Iida et al. 2012; Palen and Salzman 2002). As few social media studies been conducted via diary studies, we feel that the results of our

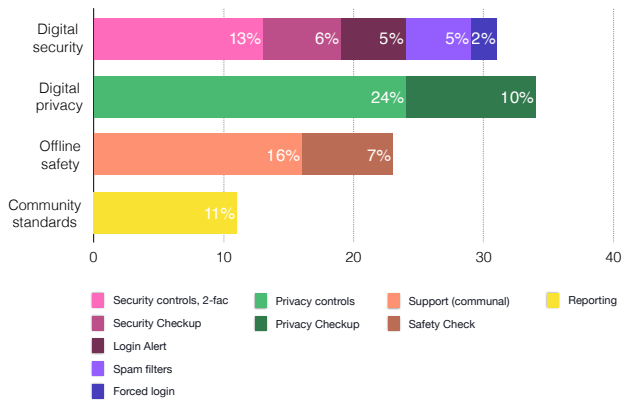

Figure 1: Proportion of experiences that made people feel safe on Facebook by coded category.

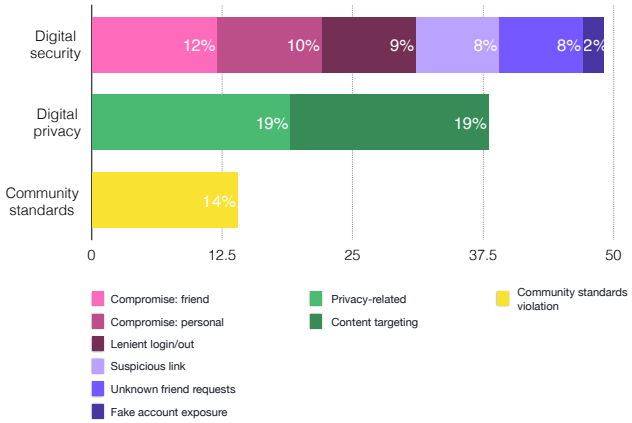

Figure 2: Proportion of experiences that made people feel threatened on Facebook by coded category.

work serve to provide another valuable set of context that would be difficult to accurately obtain through interview methodology: day-to-day experiences of safety, a topic previously unexplored, to our knowledge, through any methodology.

Additionally, as is true of qualitative work more broadly, our subject pool is not fully representative of the U.S. social-media-using population, nor can our results be highly generalized given the relatively small sample size. We did take care to recruit a diverse sample of participants[1], and a significantly larger sample than is typically recommended by research on best practices for qualitative work (Guest, Bunce, and Johnson 2006), in an effort to mitigate these limitations to the extent possible. Finally, the way in which the study was advertised or the questions in our surveys and diary experience questionnaires were phrased could have biased participants. To mitigate this, we were intentionally vague about the purpose of our study when advertising and we did not define the term "safety" for participants and worked to make survey questions as neutral as possible.

Results

We begin by detailing the results of our diary study on day-to-day safety perceptions among Facebook users. Then, we place these findings in context of other platforms by exploring participants' broader perceptions of safety and threat.

Day-to-Day Safety on Facebook

We asked participants to describe approximately three safe and three threatening experiences on Facebook over a two-week period. We also asked participants to rate the intensity of threatening experiences. Overall, participants felt safe when discovering new privacy controls, completing a Privacy or Security Checkup to audit their settings, and fulfilling additional login requirements such as two-factor authentication (2FA). Finally, and perhaps surprisingly, 76 experiences (34% of the 222 experiences we collected) involved feeling safe because of a community-building experience, such as connecting with offline community (e.g., neighbors) during a flood or being able to report harassment. (Figure 1). Participants felt threatened by targeted content, interactions with suspicious content or accounts, and violations of community standards (Figure 2).

Experiences of Safety First, we describe the participants' different day-to-day experiences of safety on Facebook.

Privacy Settings and Privacy Audits Enhance Feelings of Safety. The highest proportion of experiences related to safety (53 experiences, 24%) described the discovery of a privacy control or setting that participants could use to better tailor the audience of their posts or profile details. P23 explained: "Another thing that makes me feel secure on Facebook is my ability to tailor my friend list as well as my ability to share my posts just with specific groups of people such as close friends family or just those people from my college. This ability to tailor my posts to a specific audience overall makes me feel more safe and secure and knowing that these options are there." Similarly, another 10% of experiences involved feeling safer after seeing or being invited to go through Facebook's Privacy Checkup (Facebook f), which helps guide users through the process of setting and auditing their privacy preferences. P34 said: "Today I found a really cool measure called Privacy Checkup. You go through different questions, make sure things that are being shown are being shown to the right audiences and that certain information is everything that you want, so I was able to double-check my email wasn't being shown to the public."

Authentication Requirements Make Users Feel Safer. Relating to digital security, a total of 44 experiences (20%) described were related to authentication, including forced login after a long period of absence from a particular device (5, 2%), 2FA (6, 3%), a secondary authentication task after a password change or suspicious login attempt (22, 10%), or verification of a potentially suspicious login (11, 5%). For example, P9 described being notified about a suspicious login attempt: "I felt really safe today when I received a notification from Facebook in my email that there had been somebody signing into my account off a browser that wasn't recognized and a device that wasn't recognized...and then it was asking for this two-step security process to prove who I was and everything. I just liked the fact that Facebook was vigilant; that it pays attention to its clients, so to speak." Users'

appreciation of suspicious login notifications echoes other recent work examining suspicious login notifications and other types of security notifications more generally (Redmiles 2018b; Golla et al. 2018).

An additional 12 experiences (6%) involved Security Checkup (Facebook e), which – similar to Privacy Checkup – guides people through the set-up of two-factor authentication, login alerts, and other security features. Another 11 experiences (5%) pertained to Facebook's spam filtering mechanisms. On the latter, P2 said: "I like that Facebook is protecting my account proactively by filtering my direct messages and making sure that spam never reaches my inbox, where I can inadvertently click on something harmful or have to waste my own time deciding if something is harmful. When a message does land in my inbox, I'm more likely to trust it because I know it's been filtered through Facebook security algorithms." Similar to the suspicious login notifications, spam filtering appears to create the sense that Facebook cares about users' safety and is looking out for them.

Safety is Drawn from Community Connections Both Online and Offline. Surprisingly, the second-most frequently occurring set of safety experiences revolved around neither digital security nor privacy features, but rather around experiences on Facebook that helped people feel supported by or connected to their communities. Thirty-six (16%) experiences spoke to this type of support. P14 described an experience after a storm in her area: "I'm feeling very safe and very secure in that I'm able to connect with neighbors and my community and see who has power, who doesn't...this feature is one of my favorite parts of using Facebook. It made me feel safe and secure in the sense that I feel like I'm in the same boat with many people."

Similarly, P28 explained: "This actually happened a few moments ago... my friend updated her profile picture with the 'Help Cure Childhood Cancer' [frame] and a gold ribbon. This sort of coming together as a community and supporting a cause is something that makes me feel safe and secure. It makes me feel like I am doing all that I can to promote a cause and it really generates just a feeling of security among my Facebook friends. [Like] when the bombing happened in Paris and all these terrorist attacks happened, a lot of people including myself updated our profile pictures with sort of a frame around it, and I think that [made us feel] safe, too." These anecdotes illustrate the fluidity between online and offline safety: feelings of safety while on social media can be created through affordances that allow for users to feel better connected to and in solidarity with their offline community.

Additionally, 22 experiences (10%) pertained to Safety Check (Facebook c), a feature on Facebook in which people mark themselves as safe (or conversely, in need of help) during a natural disaster or localized crisis. Friends are subsequently notified and can offer help if needed. P24 described an experience with Safety Check during flooding that occurred at the time of the study: "It is really nice to know that friends and family are safe. I've been wondering, especially friends and family in Miami and on the West Coast in Naples, because I have not been able to get in touch with them... and then somehow they were able to check themselves safe on Facebook. It makes me feel confident about the safety of my friends and family, which is invaluable." In total, over a quarter of the experiences described involved sensations of community or physical safety, which made people feeler safer both in general and online.

Finally, 24 experiences (11%) related to feeling safer due to the ability to report content or accounts for violating Facebook's Community Standards (Facebook a). P8 said: "One thing that makes me feel really safe on Facebook is the report button. I don't use it very often. I've probably only used it maybe twice... but I love knowing that it's there. It's very extensive when you start clicking through the report button to kind of figure out your issue and make sure it goes to the right person and that it's solved appropriately." It is interesting to note that feelings of safety were generated from the *existence* of the reporting button, rather only from experiences with the *outcome* of having made a report. In some ways this sentiment similar to the feelings of safety generated from suspicious login alerts and spam filtering. The reporting button appears to generate the sense of having the platform on your side, waiting to step in; while the security features also generate a sense of partnership with the platform, who is proactively watching in the case of security affordances. The platform watching out for users appears to be a key component of safety perceptions.

Experiences of Threat Next, we describe the experiences that made people feel unsafe. These experiences fall in broadly the same categories as experiences of safety: privacy, security, and community.

Content Targeting and Unexpected Data Access Perceived As Unsafe. We found that 42 experiences (19%) that made people feel their safety was threatened pertained to content targeting, and another 42 (19%) were related to other types of privacy violations, primarily related to mismatched expectations for privacy settings. Confusion about how exactly targeted content was personalized for users led to feelings of being watched, or general concerns about how the content "got there." P16 described such an experience: "So I do actually see really great ads on Facebook for specific products that are interesting to me... but this morning I saw one and I clicked through because I was just curious in that moment how much of my information they get once I click through to one of their ads. After that, I actually began to receive solicitation product e-mails from them, and I was wondering if that was like a fluke or if indeed that type of information is being transmitted when I use Facebook." On the other hand, once one participant (P25) understood how targeting worked, they felt much safer: "a couple of hours ago I read [a] Yahoo Finance article about Facebook [content targeting]. [...] And ever since I read that article a couple hours ago I scrolled through my wall or my news feed and I was looking, I was paying more attention to the ads that are coming through to me, and I thought you know what, these are reasonable ads. They are targeting my interests. And you know, like, for the most part the information that they were advertising, it was pretty reasonable and truthful, so I felt pretty good about that."

Experiences of privacy-related threats were more varied; for example, some people felt unsafe because they unknow-

ingly gave an application permission to post on their Facebook wall, or because they did not realize that their privacy settings allowed others to see when they were attending events. P37 said: "I was shocked to keep seeing my privacy settings changed from who can see my posts. 'Friends' I set, but it says 'Public!' This reset or someone is resetting it! I need to find out why. I usually ask my daughters first. Then I will be contacting Facebook." These findings echo prior work that expectation setting is key for avoiding loss of trust and feelings of privacy violation (Rao et al. 2016; Liu et al. 2011; Netter et al. 2013). The power of mismatched expectations has been shown even in neurophysiological literature, underscoring the power and intensity of expectations to affect behavior and perception (Gaschler et al. 2014).

Suspicious People or Content and Mismatched Login Expectations Feel Unsafe. In the realm of security on Facebook, 18 experiences (8%) focused on feeling unsafe because of seeing or following a suspicious link. P31 described such an experience: "I might see either a suggested link or a link that someone else posted, and I'll see the website has, like, a weird name... like it's not some yahoo-dot-com or some well-established, well-known website. So I'm a little leery... if I click on the link, is it going to take me to a good website or is it going to be a website that gives me a pop-up to try and download that virus? I don't need to take that risk. You know, worst-case scenario." 20 experiences (9%) involved feeling unsafe because the participant was logged into Facebook when they were not otherwise expecting to be. P1 said: "I usually log on at home on my computer, and then I also log in on my laptop. Unfortunately it never logs me out. So at any given time, Facebook already has me logged in. For me, that's a security concern, because other websites, like a bank or anything that has personal information, after a short amount of time it times out and disconnects... but not with Facebook."

Participants also reported feeling their safety was threatened when they received Facebook friend requests from people they did not know or with whom they had no mutual friends (18, 8%); when they interacted with a fake account (6, 2%); when they were contacted by a friend's compromised account (27, 12%); or when recalling a time when their account had been compromised prior to the start of the study (23, 10%). Although a majority of the reported experiences did indeed occur within the two-week study timeframe, there were some exceptions. Experiences that seem to happen rarely – particularly account compromise – were sometimes reported as a distant recollection. We believe that this post-event reporting accounts for the inflated prevalence of compromise in our data. Nevertheless, these experiences – though rare – still influence participants' ongoing sense of safety on the platform, even years later.

Violations of Community Standards are Threatening. Finally, 31 reported experiences (14%) involved observing a violation of the Community Standards, which made the participant feel unsafe. For example, P12 said: "I belong to this Facebook group and for whatever reason there was a guy who started harassing the women, trying to ask them out and asking inappropriate questions. And so we reported him to the group manager, and he was removed. He just kind of kept being inappropriate, even though he was told the purpose of the group was for people to get together and go out and have fun, not just a dating area... but he wouldn't listen, and so eventually they ended up just eliminating that group and creating a new one and blocking him from it." Even though it was eventually resolved, this participant's initial experience with and observation of harassment led to sustained feelings of threatened safety.

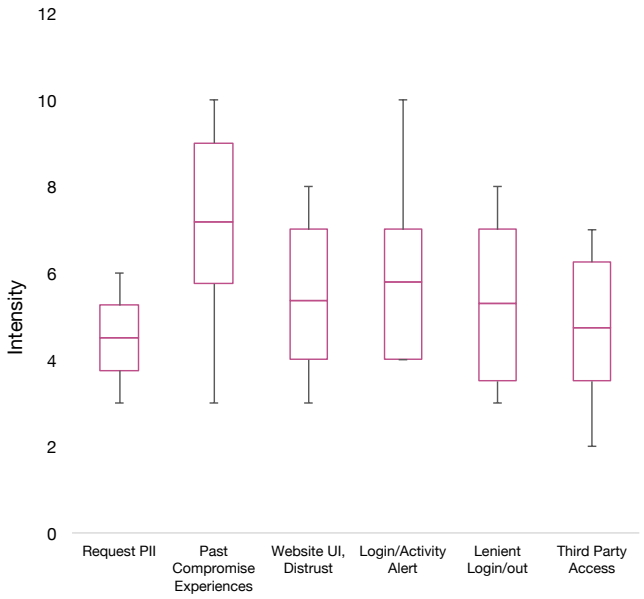

Figure 3: The mean intensity scores by experience category. The boxes show the mean and the first and third quartiles. The whiskers show the 95% confidence interval.

Community Safety Threats are Felt Most Intensely. We also asked participants how intense they felt these experiences of threatened safety on Facebook were, on a Likert scale from one to ten. Self-reported intensity of threat experiences differed significantly by threat type ($F(2,110) = 4.086$, $p = .019$). Experiences involving Community Standards – more specifically, the violation of these standards (e.g., harassment) – were significantly more intense than those involving digital privacy ($t(56) = 2.764$, $p = .008$), with a large effect size ($d = 0.893$), and those involving digital security ($t(69) = 2.056$, $p = .04$), with a medium effect size ($d = 0.635$). Interestingly, there was no significant difference between the intensity scores for security- and privacy-related experiences.

Online Safety Experiences

In the first part of our diary study, we asked participants to describe one safe and one threatening experience that they remember encountering online on any platform. Here we place participants' day-to-day safety experiences on Facebook in the context of their overall perceptions of online safety.

General Perceptions of Safety Online. Unsurprisingly, the recollected experiences that made participants feel particularly safe in general online involved many of the same attributes mentioned in the survey responses about when they felt safest on Facebook specifically. The majority (26 of 39)

of respondents described a safety experience with a bank, or with a payment- or shopping-related platform (e.g., Amazon); six respondents described an experience with email; two described an experience with a social media application; and seven described an experience with another application (e.g., a company HR website; GoDaddy.com).

The vast majority of participants (30) described an experience involving the use of multiple types of authentication, which made them feel safe. P10 said: "I have two-factor authentication set up for my account, so I cannot log in without receiving a six-digit code to my phone that I have to enter after entering my email and password. This is actually one of the few emails that I will click the link in, because I know about all four of the measures to keep my information safe." Seven participants also described a specific experience with a login or activity alert that made them feel safe; Four participants described an experience with alerts in combination with a secondary authentication request.

Additionally, four participants reported feeling safe because they were asked to update or verify their information, or in one case, to change their password. Being asked to update or verify information made P4 feel like the platform they were using had a "constant focus on security." He continued: "Just last week, I had to verify some information and they also prompted me to change my password. They do this very often, which makes me feel like my account information is well-protected." Finally, three participants recalled a specific front-end experience that made them feel more safe. P19 described signing up for a new application: "I remember I had a great first impression when I first signed up... it super easy to create my account, so I felt very secure doing so."

Experiencing Threats to Safety. Sixteen of the experiences that participants described as making them feel unsafe took place on a social media platform (Snapchat, Twitter, Facebook, and Instagram). Seven experiences took place on a banking, payment, or shopping application; five occurred on email; and ten experiences were with other applications or websites, such as news, dating, or gaming platforms.

Unsurprisingly, the most commonly described threatening experience was an experience with prior account compromise. Sixteen participants described such an experience, and one person described a friend's compromise experience that, although not experienced personally, nevertheless caused him alarm. These experiences were most heavily reported (12 of 16 participants) for banking and email platforms. Additionally, five participants described receiving an alert about a suspicious login, which made them feel threatened.

Eight participants described a poorly designed user experience or suspicious UI that made them feel threatened. P6 said: "When I first visited the site, it just didn't have a good feel to it... things were slow to load, and it was overall just not a great user experience. I'm sitting there thinking, if they don't care about that, I mean... what else do they not care about? It feels a little bit weird entering certain bits of information on a site that still had trouble just managing a loading icon. If they can't manage that, how can they manage my credit card?" Similarly, P20 described an experience with a shopping website, "I kind of got like a little scared because the authentication is not really all that great. I kept getting these repetitive things like they need to put my credit card in again or they needed to authenticate my card again...And I just felt kind of like I don't really trust this Web site."

Finally, eight participants described privacy-related situations that made them feel unsafe: four participants felt threatened when they were asked to enter personal information that they felt was unnecessary, and four participants felt threatened when a third-party application accessed information they felt was sensitive or irrelevant (e.g., a messaging application accessing their location). All but one of these participants was referring to a social media platform (not Twitter, however).

We also asked participants how intense they felt these experiences of threatened safety were, on a Likert-type scale from one to ten (Figure 3). Using a one-way ANOVA, we compared the intensity of the most commonly-described experience (past account compromise) to the other experiences. We find that participants describe past compromise as significantly more intense than experiencing requests for PII ($t(18) = 2.34$, $p = 0.03$) and significantly more intense than experiences with suspicious UX/UIs ($t(22) = 2.04$, $p = 0.049$); these differences show large effect sizes of Cohen's $d = 1.51$ and $d = 0.92$, respectively. Surprisingly, however, past compromise was not described as more intense than third-party access, remaining logged in after long periods of inactivity, or receiving alerts for suspect login attempts.

Memorable safety experiences are felt as intensely as in-the-moment experiences. Participants' mean ratings of the intensity of unsafe experiences in this first week (6.00) were not statistically different ($t = -0.017$, $df = 122$, $p > 0.05$) from the mean intensity for the day-to-day unsafe experiences recorded in weeks two and three (6.49). This suggests that in-the-moment, day-to-day events may be equally as poignant as salient experiences remembered long-term, although only the most unusual or intense events may be recalled after significant time has passed. Thus, while demand effects may be created through the diary method – and participants may feel an experience more intensely in the moment than when reported in hindsight – the similarity of the intensity of experiences of threat on a day-to-day vs. recollected basis suggests that day-to-day experiences should be considered just as important to users' overall platform experiences. Further, collecting and analyzing such day-to-day experiences, in addition to soliciting recollections of memorable events, may be useful in future work.

Discussion

Below, we present a preliminary framework for thinking about user perceptions of safety holistically, across several different dimensions (Figure 4). We discuss the need for increased research and innovation on community-related affordances, and we explore the concept of safety as a metric for measuring the "health" of a given social media platform. We conclude with a discussion of algorithmic and interface changes that may increase perceptions of safety by enhancing user control and perceived transparency.

A Framework for Thinking About Safety. We find that

Figure 4: Theoretical framework for understanding perception of safety: account is a house containing many "possessions"; safety experiences lie on one or more dimensions related to these possessions.

the distribution of safety experiences on Facebook is multifaceted, with users equally concerned with digital security; digital privacy; offline safety and a sense of community support; and the upholding of community standards.

This distribution confirms that, over the past decade, social media platforms such as Facebook have become repositories of many aspects of a person's "self": private, and potentially confidential, messages; photos and memories of childhood friends, aging offspring, and deceased loved ones; lists of friends and family members; and data that may threaten bodily safety, such as location tagging or contact information. Thus, we conceptualize safety perceptions on social media platforms as a digital "home," filled with users' valued possessions. To feel safe in a home, one must consider (1) locks, deadbolts, and security systems (in our model, this correlates to digital security); (2) closed window blinds and perimeter fencing (digital privacy); (3) contact with neighbors and loved ones to ensure a sense of community and support when in danger (offline health and safety); and (4) protection against harassing experiences, such as prank calls or vandalism (moderation and community standards). Threats to safety on social media can take many forms, evidenced by the wide variety of categories in our results; as such, a breach of safety on social media, or at least on Facebook, can be defined as any violating experience within the supposed safety and insularity of one's own home (Figure 4).

Multidimensional platforms such as social media sites present more safety challenges for users as well as for developers and designers, who must balance the needs of potentially competing dimensions (e.g., how do you balance privacy and security with the desire to share personal information for the purpose of community and relationship-building?).

On the other hand, salient threats on other services–such as online banking, for example–may be more unidimensional, as these services encompass fewer aspects of a user's self. In online banking, primarily monetary resources are at stake–and thus security may be the primary facet of safety for banking websites, with little to no emphasis on the other factors. Indeed, in week one of our diary study, those participants who described a safe or unsafe experience with an online banking site only described authentication or security concerns (or appreciation). This suggests that the use of such unidimensional platforms may make users feel safer online overall, as they have fewer "attack surfaces" and consequently may make users feel more in control of their experience. On social media platforms, however, an experience of threat to one dimension of safety could have lasting impacts on users' broader perceptions of safety across dimensions–challenging the traditional paradigm by which safety is considered, communicated, and designed for.

Within the context of social media, users may also privilege some facets of safety over others, especially with respect to the affordances of a given platform. For example, no participants mentioned a privacy threat on Twitter, while they did mention privacy violations (or support) on a variety of other social platforms (Snapchat, Instagram, Facebook). The specific affordances of the social platform – e.g., the public-by-default nature of Twitter accounts, and the lack of granular privacy controls at the tweet or reply level – may alter the significance users assign to relevant dimensions of their safety.

The findings of Cobb and Kohno (Cobb and Kohno 2017) lend further support for this hypothesis, and illustrate similarity between our social media safety taxonomy and perceptions of risk in online dating. The authors identify many dimensions of risk for online dating, some of which include a connection to offline safety and community, as well as more traditional privacy concerns, with little weight on security concerns. Future work is necessary to further define these differences in weighting between social-relevant platforms.

Localization and Community Building. We find that

over a third of the experiences that made people feel safe on Facebook were related either to a sense of community (primarily related to creating a stronger connection to offline community) or the upholding of community standards (e.g., the prevention of harassment or the ability to report content or people that violates the official community rules). Additionally, we found that the most intensely-felt unsafe experiences were those in which people were targeted by–or witnessed–someone behaving in a way that they perceived as violating Facebook's Community Standards. These results emphasize that perceptions of online safety are not exclusively related to digital privacy and security; rather, they are intertwined with users' online and offline communities, and the platform's rules and moderation practices.

Prior work on community building on social media has focused heavily on social ties (Kietzmann and others 2011; Burke, Marlow, and Lento 2010) and social capital (Ellison, Steinfield, and Lampe 2007), in addition to exploring how people create their own communities (Java et al. 2007), how community organizations use social media (Lovejoy and Saxton 2012), and the organization of community movements on social media (Juris 2012; Freelon, McIlwain, and Clark 2016). Yet community-focused social media research is typically not conducted in conversation with work on cyberbullying and harassment, intrinsically community-related experiences that our participants identify as intensely unsafe. Our results suggest that we should consider the concept of community on social media as a two-sided coin: when a community is violated or ill-protected, users may withdraw from social media due to the risk of unsafe experiences; but when community is strengthened, people feel safer and maintain the opportunity to build and capitalize from their social relationships.

Our findings underscore the importance of ongoing work to enhance various platforms' abilities to detect violations of community standards and effectively respond to harassment. Further, our findings suggest that additional focus should be given to features that support community building: for example, through hyper-localization features such as Snapchat's and Instagram's location-based and custom-group stories (Etherington 2017; Instagram). Our results imply that community-centric features will not only increase user engagement, but may also serve to increase users' overall sense of safety on the platform. Similarly, platform-supported community forums (e.g., groups on Facebook; hashtags on Twitter) may further enhance a sense of community and even offline safety (for example, groups where people can discuss current events or communicate safety statuses after natural disasters).

Finally, prior work has explored how these same community and movement-building features can be leveraged to amplify or increase harassment and other abusive behaviors (Flores-Saviaga, Keegan, and Savage 2018). In sum, affordances that allow users to build, enhance, or maintain communities should be designed with consideration for users' safety perceptions and preferences. Such affordances should also be viewed through multiple lenses: while these features can help to engage and entertain users *and* make them feel safer, violations can threaten users' sense of safety with an intensity that may rival or even exceed traditional privacy and security threats. Future work should explore the design of community-building tools with this context in mind, to minimize the potential for abuse and resulting threats to users.

Metrics for Community and Social Media Health. Social media companies are increasingly evaluating the success of their platforms through the framework of community health. While a significant body of work has measured community health on social media sites based on negative metrics such as harassment prevalence (Twitter 2018), our work suggests it may be prudent to measure health using both negative *and* positive metrics, such as users' perception of personal safety and sense of community (Duggan et al. 2017; Ybarra and Mitchell 2008). Such metrics can be provided as feedback to moderators of sub-communities (e.g., Facebook groups) to help empower them to strengthen new and existing bonds, or used to filter and rank content in individual users' feeds.

Further, our results suggest that designers should consider how well a given online community supports *offline* community-building as a component of overall platform health. For example, while prior work has explored how people represent or misrepresent real-world events on social media and how news of disasters spreads (Qu et al. 2011; Kim et al. 2012), relatively little work – especially quantitatively – has explored how users create support for each other around these events, particularly *in the moment* rather than after the fact. Features that enable offline support or enhance existing offline communities contribute positively to social media users' overall sense of safety and, consequently, their engagement.

Drawing from our conceptual framework of safety, we suggest that social media health should be viewed not only through a community lens, but according to users' overall perception of their safety on a given platform. Future work establishing how "safety" could be easily evaluated and tracked over time (e.g., through a regularly-occurring survey) would allow platforms to better contextualize and enhance other goals–for example, adjusting algorithmic feeds or targeted advertising to improve perceptions of overall platform safety. Such a descriptive, bottom-up approach – in which user perceptions drive algorithmic changes – has been suggested in other algorithmically-governed contexts (e.g., (Grgic-Hlaca et al. 2018b; 2018a) and could be useful in other contexts: for example, which community groups are suggested to users; what security features are added or prioritized; or for making recommendations to policy makers regarding the regulation of data collection and online harassment.

That said, we must be careful to balance creating "safety theater" (Soghoian 2011) – or the deceptive generation of an inappropriate sense of safety, which may incline users to act without fear of consequences (when consequences may indeed exist) – with the need to foster self-efficacy, particularly through helping users' perceptions that match reality (Schneier 2007).

Increasing Perceived Control through Transparency. Finally, to specifically address the safety dimensions of privacy and security, we suggest additional focus on perceived control. Prior work has found that a primary predictor of risk

perception is the degree to which an individual feels that they have sufficient efficacy to control a present threat (Shin 2010; Witte 1998; 1994). We find that while nearly a quarter of reported experiences that made participants feel safe on Facebook involved the ability to control their privacy, only 6% of participants described experiences with actively enhancing *security*. This lack of security control translates into increased experiences of threat, as more participants described feeling unsafe on Facebook due to a security-related event (49%) rather than a privacy-related event (38%). In some of these cases, participants expressed a desire for already-existing security features such as 2FA (Song 2011) or reported more generally that they found the platform's security settings difficult to find. Thus, while security can quickly become burdensome or overwhelming for users, our results point to a need to make security options and controls more visible and accessible through educational campaigns (Slatery) and proactive surfacing of relevant features.

Similarly, 10% of our participants described feeling unsafe on Facebook due to a lack of understanding around content personalization. When people do not understand how content is being tailored to them, they often report believing that someone is inappropriately tracking them, reading their messages, or accessing their account details. This notion that content is personalized based on demographics or other inferred personal characteristics is of much greater concern to users (Plane et al. 2017) than the thought of targeting based on objective, aggregated behavior such as clicking behaviors. To this end, we recommend proactive education to increase user understanding around content targeting practices on social media sites (Facebook b), as well as enabling users to quickly view and manage their advertising preferences (Facebook d).

Beyond privacy, transparency has also been shown to be important for democratic community-building on social media platforms. Blackwell et al. argue that visible disclosure of the criteria and process by which online harassment is identified, categorized, and addressed can reduce feelings of exclusion and enhance users' overall sense of community (Blackwell et al.). Relatedly, Crawford and Gillespie raise the question of whether flagging and removing potentially violating content should be made more visible to users–for example, by showing that a piece of content was removed and why, or by allowing users to appeal or debate the validity of the flag (Crawford and Gillespie 2016). We can draw many parallels from these ideas to advertising transparency. The first, providing users with explanations with why they see advertisements, is already in place on some platforms. Future work may wish to explore the effects of such explanations on users' general perceptions of safety, as well as exploring the effect of taking a more social and interactive approach. For example, would allowing users the opportunity to discuss the advertisements they are shown–or the personalizations they experience–help them gain knowledge, self-efficacy, and ultimately a greater sense of control and safety that would enhance their online experiences? Our results suggest yes.

Summary

We conducted a diary study (n=39) to better understand what makes people feel safe on Facebook. We contribute a multidimensional taxonomy of the experiences and affordances which affect people's perceptions of safety on Facebook, and we offer a theoretical framework for considering these four dimensions of online safety: digital security; digital privacy; offline safety; and community standards. Our results underscore the importance of fostering a sense of community online, as community-related experiences have a significant impact on safety perceptions–which in turn influence users' self-efficacy, and may thus bolster or inhibit their adoption of controls that support safety. Ultimately, we suggest viewing online safety through a multifaceted lens–that is, moving beyond considerations for just privacy or security in isolation, but rather considering and addressing a more holistic concept of safety.

References

AdWeek. How Much Time Will the Average Person Spend on Social Media During Their Life? (Infographic) Adweek.

Ashktorab, Z., and Vitak, J. 2016. Designing cyberbullying mitigation and prevention solutions through participatory design with teenagers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3895–3905. ACM.

Ashktorab, Z. 2018. the continuum of harm taxonomy of cyberbullying mitigation and prevention. In *Online Harassment*. Springer. 211–227.

Bauer, L., and et al. 2013. The post anachronism: The temporal dimension of facebook privacy. In *WPES*. ACM.

Blackwell, L.; Dimond, J.; Schoenbeck, S.; and Lampe, C. Classification and its consequences for online harassment: Design insights from heartmob.

Blackwell, L.; Chen, T.; Schoenebeck, S.; and Lampe, C. 2018. When online harassment is perceived as justified. *ICWSM*.

Bodford, J. E. 2017. *Blurring Safety Between Online and Offline Worlds: Archival, Correlational, and Experimental Evidence of Generalized Threat in the Digital Age*. Ph.D. Dissertation, Arizona State University.

Boyd, D. 2014. *It's complicated: The social lives of networked teens*. Yale University Press.

Braun, V., and Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative research in psychology*.

Burke, M.; Kraut, R.; and Marlow, C. 2011. Social capital on facebook: Differentiating uses and users. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 571–580. ACM.

Burke, M.; Marlow, C.; and Lento, T. 2010. Social network activity and social well-being. In *SIGCHI*. ACM.

Camp, L. J. 2009. Mental models of privacy and security. *IEEE Technology and society magazine*.

Carter, S., and Mankoff, J. 2005. When participants do the capturing: the role of media in diary studies. In *CHI*. ACM.

Center, P. R. 2017. Social media fact sheet.

Cobb, C., and Kohno, T. 2017. How public is my private life?: Privacy in online dating. In *WWW*.

Crawford, K., and Gillespie, T. 2016. What is a flag for? social media reporting tools and the vocabulary of complaint. *New Media & Society*.

Druin, A., and et al. 2010. Children's roles using keyword search interfaces at home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM.

dScout.

Duggan, M.; Lee, R.; Smith, A.; Funk, C.; Lenhart, A.; and Madden, M. 2017. Online harassment 2017. *The Pew Research Center.(11 July 2017). Retrieved September* 8:2017.

Ellison, N. B.; Vitak, J.; Steinfield, C.; Gray, R.; and Lampe, C. 2011. Negotiating privacy concerns and social capital needs in a social media environment. In *Privacy online*.

Ellison, N. B.; Steinfield, C.; and Lampe, C. 2007. The benefits of facebook friends: social capital and college students use of online social network sites. *Journal of Computer-Mediated Communication* 12(4):1143–1168.

Etherington, D. 2017. Snapchat now lets you create custom stories for groups of friends and family.

Facebook. Community standards.

Facebook. How ads work on facebook.

Facebook. I'm seeing stories about my friends being marked safe through safety check in my news feed. what does this mean?

Facebook. What are my ad preferences and how can i adjust them?

Facebook. What's security checkup and how do i start it?

Facebook. What's the privacy checkup and how can i find it?

Flores-Saviaga, C.; Keegan, B. C.; and Savage, S. 2018. Mobilizing the trump train: Understanding collective action in a political trolling community. *ICWSM*.

Freelon, D.; McIlwain, C. D.; and Clark, M. D. 2016. Beyond the hashtags:# ferguson,# blacklivesmatter, and the online struggle for offline justice.

Furnell, S.; Bryant, P.; and Phippen, A. D. 2007. Assessing the security perceptions of personal internet users. *Computers & Security*.

Gaschler, R.; Schwager, S.; Umbach, V.; Frensch, P.; and Schubert, T. 2014. Expectation mismatch: differences between self-generated and cue-induced expectations. *Neuroscience & Biobehavioral Reviews* 46:139–157.

Gibbs, J. L.; Ellison, N. B.; and Lai, C.-H. 2011. First comes love, then comes google: An investigation of uncertainty reduction strategies and self-disclosure in online dating. *Communication Research*.

Gl´uer, M., and Lohaus, A. Frequency of victimization experiences and well-being among online, offline, and combined victims on social online network sites of german children and adolescents. *Frontiers in public health*.

Golla, M.; Wei, M.; Hainline, J.; Filipe, L.; Drmuth, M.; Redmiles, E. M.; and Ur, B. 2018. "what was that site doing with my facebook password?" designing password-reuse notifications. *Proceedings of the 25th ACM Conference on Computer and Communications Security (CCS)*.

Grgic-Hlaca, N.; Redmiles, E. M.; Gummadi, K. P.; and Weller, A. 2018a. Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, 903–912. International World Wide Web Conferences Steering Committee.

Grgic-Hlaca, N.; Zafar, M. B.; Gummadi, K. P.; and Weller, A. 2018b. Beyond distributive fairness in algorithmic decision making: Feature selection for procedurally fair learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA*.

Guest, G.; Bunce, A.; and Johnson, L. 2006. How many interviews are enough? an experiment with data saturation and variability. *Field methods* 18(1):59–82.

Hargittai, E., and Litt, E. 2011. The tweet smell of celebrity success: Explaining variation in twitter adoption among a diverse group of young adults. *New media & society* 13(5):824–842.

Hayes, A. F., and Krippendorff, K. 2007. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*.

Holmes, K. M., and O'loughlin, N. 2014. The experiences of people with learning disabilities on social networking sites. *British Journal of Learning Disabilities*.

Iida, M.; Shrout, P. E.; Laurenceau, J.-P.; and Bolger, N. 2012. Using diary methods in psychological research.

Instagram. Hashtag and location stories on explore.

Java, A.; Song, X.; Finin, T.; and Tseng, B. 2007. Why we twitter: understanding microblogging usage and communities. In *WebKDD*. ACM.

Joyce, M. 2013. Picking the best intercoder reliability statistic for your digital activism content analysis.

Juris, J. S. 2012. Reflections on# occupy everywhere: Social media, public space, and emerging logics of aggregation. *American Ethnologist*.

Kietzmann, J. H., et al. 2011. Social media? get serious! understanding the functional building blocks of social media. *Business horizons*.

Kim, M.; Xie, L.; Christen, P.; et al. 2012. Event diffusion patterns in social media. In *ICWSM*.

Kim, J.; LaRose, R.; and Peng, W. 2009. Loneliness as the cause and the effect of problematic internet use: The relationship between internet use and psychological well-being. *CyberPsychology & Behavior*.

Kross, E., and et al. 2013. Facebook use predicts declines in subjective well-being in young adults. *PloS one*.

Lee, D.; Larose, R.; and Rifon, N. 2008. Keeping our network safe: a model of online protection behaviour. *Behaviour & Information Technology*.

Litt, E. 2013. Understanding social network site users? privacy tool use. *Computers in Human Behavior*.

Liu, Y.; Gummadi, K. P.; Krishnamurthy, B.; and Mislove, A. 2011. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, 61–70. ACM.

Lovejoy, K., and Saxton, G. D. 2012. Information, community, and action: How nonprofit organizations use social media. *Journal of Computer-Mediated Communication*.

Madden, M., and et al. 2013. Teens, social media, and privacy. *Pew Research Center*.

McNally, B., and et al. 2017. Gains from participatory design team membership as perceived by child alumni and their parents. In *CHI*. ACM.

Morris, M. R.; Teevan, J.; and Panovich, K. 2010. A comparison of information seeking using search engines and social networks. *ICWSM* 10:23–26.

Netter, M.; Riesner, M.; Weber, M.; and Pernul, G. 2013. Privacy settings in online social networks–preferences, perception, and reality. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, 3219–3228. IEEE.

Page, X.; Kobsa, A.; and Knijnenburg, B. P. 2012. Don't disturb my circles! boundary preservation is at the center of location-sharing concerns. In *ICWSM*.

Palen, L., and Salzman, M. 2002. Voice-mail diary studies for naturalistic data capture under mobile conditions. In *CSCW*. ACM.

Plane, A. C.; Redmiles, E. M.; Mazurek, M. L.; and Tschantz, M. C. 2017. Exploring user perceptions of discrimination in online targeted advertising. In *USENIX Sec*.

Qu, Y.; Huang, C.; Zhang, P.; and Zhang, J. 2011. Microblogging after a major disaster in china: a case study of the 2010 yushu earthquake. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 25–34. ACM.

Rao, A.; Schaub, F.; Sadeh, N.; Acquisti, A.; and Kang, R. 2016. Expecting the unexpected: Understanding mismatched privacy expectations online. In *Symposium on Usable Privacy and Security (SOUPS)*, volume 4, 2.

Redmiles, E. M.; Acar, Y.; Fahl, S.; and Mazurek, M. L. 2017. A summary of survey methodology best practices for security and privacy researchers. Technical report.

Redmiles, E. M.; Chachra, N.; and Waismeyer, B. 2018. Examining the demand for spam: Who clicks? In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 212. ACM.

Redmiles, E. M. 2018a. Net benefits: Digital inequities in social capital, privacy preservation, and digital parenting practices of us social media users. *ICWSM*.

Redmiles, E. M. 2018b. "should i worry" a cross-cultural examination of account security incident response. *arXiv preprint arXiv:1808.08177*.

Schneier, B. 2007. In praise of security theater. *Schneier on Security* 25.

Scollon, C. N.; Prieto, C.-K.; and Diener, E. 2009. Experience sampling: promises and pitfalls, strength and weaknesses. In *Assessing well-being*. Springer. 157–180.

Shin, D.-H. 2010. The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption. *Interacting with computers*.

Slatery, B. Learn how to stay vigilant and report scams.

Smith, A., and Anderson, M. 2018. Social media use in 2018.

Soghoian, C. 2011. An end to privacy theater: Exposing and discouraging corporate disclosure of user data to the government. *Minn. JL Sci. & Tech.* 12:191.

Song, A. 2011. Introducing login approvals.

Stenson, C.; Balcells, A.; and Chen, M. 2015. Burning up privacy on tinder. *SOUPS (Posters)*.

Strater, K., and Lipford, H. R. 2008. Strategies and struggles with privacy in an online social networking community. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 1*, 111–119. British Computer Society.

Tausczik, Y. R.; Wang, P.; and Choi, J. 2017. Which size matters? effects of crowd size on solution quality in big data q&a communities. In *ICWSM*.

Twitter. 2018. Twitter health metrics proposal submission.

Udo, G. J. 2001. Privacy and security concerns as major barriers for e-commerce: a survey study. *Information Management & Computer Security* 9(4):165–174.

Winnick, M. 2012. dscout. In *Ethnographic Praxis in Industry Conference Proceedings*.

Wisniewski, P.; Knijnenburg, B. P.; and Lipford, H. R. 2014. Profiling facebook users? privacy behaviors. In *SOUPS2014 Workshop on Privacy Personas and Segmentation*.

Witte, K. 1994. Fear control and danger control: A test of the extended parallel process model (eppm). *Communications Monographs*.

Witte, K. 1998. Fear as motivator, fear as inhibitor: Using the extended parallel process model to explain fear appeal successes and failures.

Ybarra, M. L., and Mitchell, K. J. 2008. How risky are social networking sites? a comparison of places online where youth sexual solicitation and harassment occurs. *Pediatrics* 121(2):e350–e357.

Zhang, H.; De Choudhury, M.; and Grudin, J. 2014. Creepy but inevitable?: the evolution of social networking. In *CSCW*. ACM.