

“It’s easier than causing confrontation”: Sanctioning Strategies to Maintain Social Norms and Privacy on Social Media

YASMEEN RASHIDI, APU KAPADIA, CHRISTENA NIPPERT-ENG, and NORMAN MAKOTO SU, Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington, USA

Sanctions play an essential role in enforcing and sustaining social norms. On social networking sites (SNS), sanctions allow individuals to shape community norms on appropriate privacy respecting behaviors. Existing theories of privacy assume the use of such sanctions but do not examine the extent and effectiveness of sanctioning behaviors. We conducted a qualitative interview study of young adults (N=23), and extend research on collective boundary regulation by studying sanctions in the context of popular SNS. Through a systematization of sanctioning strategies, we find that young adults prefer to use indirect and invisible sanctions to preserve strong-tie relationships. Such sanctions are not always effective in helping the violator understand the nature of their normative violation. We offer suggestions on supporting online sanctioning that make norms more visible and signal violations in ways that avoid direct confrontation to reduce the risk of harming on-going social relationships.

CCS Concepts: • **Security and privacy** → **Social aspects of security and privacy**; **Usability in security and privacy**.

Additional Key Words and Phrases: norms; sanctions; communication privacy management; boundary regulation; social networking sites

ACM Reference Format:

Yasmeen Rashidi, Apu Kapadia, Christena Nippert-Eng, and Norman Makoto Su. 2020. “It’s easier than causing confrontation”: Sanctioning Strategies to Maintain Social Norms and Privacy on Social Media. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 23 (May 2020), 25 pages. <https://doi.org/10.1145/3392827>

1 INTRODUCTION

Social norms guide our everyday behavior and enable groups of people to interact smoothly [68]. As they use different technologies to communicate with friends, family members, and coworkers, for instance, groups of college students with a shared sense of identity collectively create ‘idioms of practice,’ i.e., sets of normative expectations around the ways individuals should and should not use these technologies with each other [33]. Norms like these are established when “any departure of real behavior from the norm is followed by some punishment” [39, p. 123]. Sanctions are then used to signal, enforce, and sustain social norms [58]. Sanctions allow a group to exert ‘social control’ over each other – by formal or informal and verbal or nonverbal means – and express disapproval of members’ counter-normative behaviors [19, 27].

Authors’ address: Yasmeen Rashidi, yrashidi@indiana.edu; Apu Kapadia, kapadia@indiana.edu; Christena Nippert-Eng, cnippert@indiana.edu; Norman Makoto Su, normsu@indiana.edu, Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington, 700 N Woodlawn Ave, Bloomington, Indiana, USA, 47408.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Association for Computing Machinery.

2573-0142/2020/5-ART23 \$15.00

<https://doi.org/10.1145/3392827>

We seek to understand how young adults – who have the highest adoption rate and frequency of use with the greatest variety of social media platforms today [85] – maintain collective norms of content sharing on social media. We are especially interested in how they sanction violations of these norms in the context of “inappropriate” content sharing and privacy violations. In particular, we seek to understand the use of ‘negative’ sanctions, which are social responses that “punish or otherwise discourage violations of social norms and symbolically reinforce the culture’s values and morals” [68, p. 90]. The tension that arises from the use (or non-use) of negative sanctions can provide a rich opportunity for design insights aimed to improve a community’s enforcement or communication of its expected norms on social media.

Social Networking Sites (SNS) are now an indispensable part of young adults’ social lives for interacting with others and sharing content online. Anything these individuals reveal about themselves (e.g., photos, music, links, and short text-based updates) or others (e.g., by tagging friends in photos or locations) can, therefore, potentially affect the privacy [73, 78], self-presentation [36, 94], and reputation [87] of many parties [25, 74, 99]. This is partly because, in today’s networked world, once personal information is disclosed, “it moves into a collective domain where the information is no longer under the sole control of the individual” [74, p. 19]. For example, when a photo is shared, the responsibility for maintaining the privacy of its content is shared by not only the photographer and the subjects [74, 92, 104] but also by the viewers [74]. In the context of boundary regulation theory [74], all parties are considered ‘co-owners’ of such shared information.

Given the significance of the collective’s influence on the privacy of shared content, an appropriate theoretical lens through which to approach this research is Petronio’s [74] seminal theory of Communication Privacy Management (CPM). Adopted in various studies of managing private information disclosure online, CPM builds on Altman’s theory of privacy [1] by explaining how individuals dialectically and dynamically manage ‘co-owned’ information and social boundaries together [74]. The main thrust of CPM theory is that privacy ‘rules’ are developed through interpersonal boundary management and collaborative negotiations where factors such as cultural expectations (e.g., social norms [68]) and social contexts play an essential role. ‘Norms of appropriateness’ determine what information is appropriate to reveal, to whom, and in what way, while ‘norms of distribution’ restrict the flow of information within and across contexts [70]. When individuals violate privacy rules regarding appropriateness or distribution, a ‘negative sanction’ may be applied to the violator in order to reinforce the existence and importance of following these norms. This is especially important when implicit rules are violated, as a negative sanction takes a rule that was only implied by the collective and makes it more apparent [74]. While a significant amount of research has focused on how co-owners of private information manage their disclosure collaboratively on SNS [44, 54, 59, 67, 78, 88, 100], the role of the audience in this process is still unclear. Our study extends research on collective boundary regulation in SNS by focusing on the role of “viewers” as co-owners of information to which the poster/owner has granted them access.

Researchers have examined social norms and violations on SNS (such as Facebook) in regard to the evolution and detection of norms [17, 40, 50, 64], and how these norms affect privacy [24, 91, 103] and self-presentation [97]. However, as previous authors acknowledge, we do not yet have a systematic understanding of how different norms on SNS are enforced, how violations are detected, and how sanctions are applied in different social contexts [40, 53]. Additionally, not much is known about how ‘invisible interactions’ (e.g., offline confrontations or choosing to ignore inappropriate content) affect people’s use of SNS or the quality of their relationships with others [29]. Understanding the social norms of appropriateness surrounding shared online content could help individuals avoid violations of these norms as well as invasions of co-owners’ privacy [59], which might affect people’s self-presentation online [94], or expose them to sanctions by the viewers [64]. A better understanding of communal expectations regarding which information should be shared, managed,

and withheld, and how norms and sanctions are enacted as a part of everyday social life, can enhance the privacy and the usability of these interactive systems [26].

In this paper, we contribute to the scholarly research about these communal practices by: (1) investigating norms and sanctions across the multiple social platforms used by our participants (i.e., Facebook, Snapchat, Twitter, and Instagram) and (2) systematically analyzing the strategies used by these young adults to sanction their peers whenever violations of social norms occur. To do so, we conducted semi-structured interviews with 23 college students. We focused on three research questions: (1) *What kinds of behaviors and content sharing do college students identify as inappropriate on social media?* (2) *How do students sanction each other when someone violates these social norms of appropriateness?* and (3) *How might designs draw from these sanctioning strategies to better support college students’ goals?*

We find that the negative sanctions used by these students fall under three dimensions: *who* performs the sanction (a person or a group), *where* they sanction (on-site or off-site), and *how* apparent the sanction is to the violator (visible or invisible). Overall, college students approach sanctioning *discreetly*, preferring to avoid or internalize conflict (e.g., by gossiping about or downplaying the violation) and – when needed – to sanction others in indirect or passive ways, sometimes collaboratively. Our findings show that current social media designs are not helpful in highlighting norms while also maintaining the sanctioners’ goals of avoiding conflict and keeping the peace. We argue for designing sanctioning tools that make community norms more *visible* and *signal* norm violations of sharing. Such approaches do not stifle change, and also support norm-violations that may be seen as necessary by certain community members. These tools, unlike those that enforce norms or call for conformity, should further alleviate the burden on viewers to confront the violator while saving face and preventing public embarrassment for both the poster and the viewers on SNS. Making such dialog more visible may thus enable both a healthy enforcement and evolution of community norms.

2 RELATED WORK

In this section, we describe work related to privacy management as well as social norms and sanctions in social networking sites.

2.1 Privacy as Collective Boundary Regulation

Privacy is inherently a social process; individuals seek a balance between the privacy and disclosure of their personal information to manage their relationships with others [69]. Within social networking sites – where the processes of information disclosure and dissemination are affected by the audience’s co-existence, social norms, and technical functionality – context plays an essential role in determining privacy practices [63]. Nissenbaum [70] unpacks privacy through the notion of ‘contextual integrity’ (CI), which is useful for understanding the cultural expectations around the transfer of personal information [70]. Understanding the social norms governing information flow is a crucial starting point for understanding the appropriateness of information sharing in a given context. Contextual integrity is preserved (i.e., there is no privacy violation) whenever information sharing norms are upheld. According to Marwick and boyd, achieving networked privacy today requires an “ongoing negotiation of the contexts in a networked ecosystem in which contexts regularly blur and collapse” [63, p. 13]. This negotiating and managing of one’s personal information, privacy, and identity is no longer an individualistic process; it is increasingly seen as a collective one [32, 44, 54, 59, 63, 67, 70, 73, 74, 78, 88, 100]. Petronio labeled this dialectic, dynamic process of negotiating privacy boundaries ‘Communication Privacy Management (CPM)’ [74].

According to CPM, once information is ‘disclosed’ and shared with others, it moves into a collective domain of ‘co-owners’ where the information is no longer under the sole control of the

individual. To manage what are now collective boundaries, individuals create regulation rules that are impacted by cultural expectations and social contexts, either learning the rules or negotiating to create new ones. When negotiation fails, boundary ‘turbulence’ is likely. In response to such violations, people apply sanctions to address the breach of privacy and ask the violator to comply with the rule.

A growing body of privacy research has studied the mechanisms users have developed to manage privacy and collective boundaries on SNS (i.e., co-managing shared content) [7, 15, 23, 54, 55, 63, 78, 90, 91, 102]. Lampinen et al. [54] created a three dimensional framework of strategies for managing private versus public disclosures on SNS: ‘mental and behavioral’, ‘preventive and corrective’, and ‘individual and collective’. In this framework, people applied a set of ‘mental’ strategies (e.g., trusting others to be considerate of one’s boundary regulation) or ‘behavioral’ strategies (e.g., using deliberate wording and tone in posts). These strategies can be further divided into ‘preventive’ (e.g., avoiding sharing content that could be problematic) or ‘corrective’ (e.g., deleting content). People apply these strategies ‘individually’ (e.g., adjusting privacy settings to disable disclosure) or ‘collaboratively’ (e.g., negotiating and agreeing on a set of rules). Another way to categorize these strategies is by distinguishing between ‘online’ strategies, supported by the user interfaces of SNS, and ‘offline’ strategies, used by individuals when they cannot find a technical tool within a given platform to support their needs [7, 54, 78, 102]. Rashidi et al. [78], for example, found that when friends violate each other’s privacy, they tend to confront each other offline (i.e., face-to-face) to directly negotiate the violation. In addition, viewers remain vigilant and watch their friends’ backs to warn them about any violations of their privacy on SNS.

Most of the previous research on collaborative boundary regulation has focused on the privacy practices of the original owner of the shared information and the other co-owners of the shared content (see for example [44, 54, 59, 63, 88]). However, to the best of our knowledge, previous work has rarely focused on the role of viewers as co-owners of the shared content, including how viewers manage privacy when other people on social media violate privacy rules. In this work, we shed light on the role of viewers in the collaborative process of managing privacy on SNS. We also use the literature on collective boundary regulation and privacy management strategies to extend our understanding of sanctions, in part by systematizing the kinds of sanctioning strategies used by individuals on SNS. This understanding will allow us to explore designs that provide SNS-users with better tools to manage privacy online.

2.2 Social Norms and Sanctions

Sociologists see social norms as “culturally defined rules of conduct. They specify what people should do and how they should pursue values. They tell us what is proper or necessary behavior within particular roles, groups, organizations, and institutions” [68, p. 31]. In this work, we approach social norms on SNS as *implicit (informal) norms* that are unspoken, not formally written or recorded, but generally understood by a social group [14, 68, 80]. Social norms are dynamic in their nature; they change over time [34], they are cultural and context dependent [34, 76], and they influence how one behaves while also being shaped by members’ behaviors [76]. Norms provide us with a general framework of expectations but never tell us exactly how to act. Rather, people learn social norms by interacting and observing others, as well as by being sanctioned for violating these norms [34, 56, 82]. Therefore, social norms – especially if individuals do not internalize them [41] – are usually accompanied by sanctions [34, 39, 68], i.e., “penalties and rewards for conduct concerning a norm” [80, p. 73]. A sanction can be “any reaction from others to the behavior of an individual or group that is meant to ensure that the person or group complies with a given norm” [34, p. 149]. Sanctions can be positive (i.e., reward for conformity) or negative (i.e., punishment for nonconformity) [34, 80]. They can also be formal (e.g., imposed by an authoritative force, such as

Facebook’s administrators or law enforcement) or informal (e.g., imposed by users or ordinary people in the group) [34, 80]. In this work, we use the word “sanction” to refer to informal negative sanctions imposed by peer viewers.

When a sanction is observed, it is clear evidence that a norm does exist [58, 71]. People are more likely to conform to a social norm if they perceive that noncompliance will be accompanied by negative sanctions and if this norm is widespread among their peers [10, 56]. Imposing sanctions is costly, not just for the violator, but for others who are responsible for applying sanctions [27] as well since they incur “the risk of retaliation or at least the potential loss of a relationship, the loss of time or money, emotional tensions, and so forth” [30, p. 2]. Therefore, a person who sanctions a norm violator needs to have a strong motivation to carry out the punishment [75]. However, individuals might bear the costs of sanctioning if the potential benefits of their actions exceed those costs [18].

In online communities, norms also have an important role in guiding acceptable behaviors [57]. Implicit social norms (as opposed to explicit norms) put pressure on online community members to identify and obey them, and they incentivize individuals to forgo questionable behavior – especially when violating them will result in social sanctions (e.g., disparaging remarks or loss of reputation) [14, 32, 84]. Building on Ostrom’s framework [72] of the collective governance of shared ‘common-pool’ resources, Garg et al. [32] argued that designs for community governance of privacy require encouraging dialogue between community members, enabling transparency of information flows, and supporting enforcement of community norms. Social norms and sanctions have been studied in the field of online communities through technology specific studies of, for instance, social networking sites (i.e., Facebook) [40, 50, 64, 79, 97], online dating services [28], live streaming platforms [82], social voting sites [17] and other online communities [14, 66, 84, 105].

Due to the popularity of Facebook among college students, norms and violations on Facebook have been well studied [40, 50, 64, 79, 97, 103]. Hooper and Kalidas [40] provided a list of unacceptable behaviors and content on Facebook that emerged among college students, including offensive, unprofessional, or embarrassing posts and breaching others’ privacy. Heated interactions in public (i.e., on Facebook ‘walls’ or through comments) were also seen as inappropriate and a norm violation on Facebook [64]. To explain how people responded to unexpected behaviors on Facebook, McLaughlin and Vitak [64] studied the evolution of norms and their violations through the lens of Expectancy Violation Theory (EVT) [12, 13]. EVT states that deviation from expected behaviors always brings more attention to the violator and to the correct interpretation of the violation. The viewers’ reactions toward the unexpected behaviors (whether negative or positive) usually depend on their relationships with the violator and the context (e.g., environmental constraints and privacy) of the situation [13]. McLaughlin and Vitak [64] further found that when norm violations occur, the closeness of the relationship plays an essential role in determining the sanction. With friends, individuals were more likely to confront the violators whereas with acquaintances the severity of the violation played a role in determining the right sanction. However, McLaughlin and Vitak’s work did not discuss how people react to minor norm violations (violations that do not directly impact their privacy or self-presentation goals) from their friends. Wolfer [103] also studied college student perception of Facebook drinking posts and found that many participants considered posts depicting underage drinking and drunken vomiting as inappropriate; however, the vast majority did not sanction such posts by leaving negative comments, preferring to ignore such posts.

Taking into account the strong bond between norms and sanctions, our position is that eliciting stories about sanctioning is an effective means of identifying norms, since sanctions manifest when a norm has been violated. We are primarily interested in the different strategies young adults use to sanction inappropriate behaviors and content. In addition to understanding the sanctions and

norms, themselves, we also aim to understand whether sanctions succeed in making violators understand the social norms they violated, so that future violations might be reduced.

3 METHODS

We conducted semi-structured interviews with 23 college students over a five-week period on a large college campus. Students were recruited through flyers placed in common areas and online university classifieds. Our eligibility criteria for participation was that a participant must be at least 18 years old and a frequent or active user of Facebook and Snapchat. Participants completed an informed consent form at the beginning of each interview. Our university human subjects ethics board approved this study.

The interviews lasted 36–85 minutes ($M = 56$ min, $SD = 11.65$ min), and each participant was compensated \$15 USD upon completion of the study. Of the 23 participants, 16 identified as female and 7 identified as male. Most of the participants (18) were undergraduate students, five were graduate students (all aged 18–26). They came from diverse fields of study, and represented different ethnicities (White (11), Asian (7), African American (6), Caucasian (5), and European (1)). All participants reported they had been using Snapchat and Facebook. Most (20) were using Instagram, a majority (15) were using Twitter, and a few (<3) were using LinkedIn, Tinder, Tumblr, and WhatsApp. The majority (18) reported posting content on their social media accounts frequently or occasionally.

All the interviews were conducted in person. We employed critical incident technique [20]; once participants told us stories/incidents, we probed for specific details, allowing participants to control the narrative and help us understand what occurred from their perspectives. The audio of the interviews was recorded. After the first nine interviews were complete, we transcribed, anonymized, and analyzed the collected data. First, one of the researchers analyzed the transcripts using an iterative coding process involving both open and axial coding [89] to create an initial code book. Then, all researchers met multiple times to discuss the identified themes. These discussions led to an update of the semi-structured interview protocol focused around select themes that we felt were underdeveloped and deserved more probing. This updated, focused protocol was then used with the remaining 14 participants. Researchers met again and analyzed the collected data using the aforementioned process. We reached thematic saturation after the first 16 interviews, and the remaining seven interviews confirmed our identified recurring themes. Memoing, incorporating our code book, was done throughout the process. These memos revolved around codes such as “Where did the violation occur (e.g., Snapchat, Twitter, or Facebook)?”, “Sanctioner relationship with the violator (e.g., friend, family member, or acquaintance),” and “How did sanctioner sanction (e.g., on-site, off-site, collaboratively, or/and individually)?” which form the basis of our paper’s findings and discussion.

Our initial semi-structured interview protocol investigated the sanctioning behaviors people use when someone violates their expectations about what or how content should be shared on certain social media. The final protocol emphasized our interest in the similarities and differences of social norms on different social media sites, people’s sanctioning reactions associated with the violation of these norms, and participants’ evolving attitudes and behaviors around content sharing and collective sanctioning on social media. Each participant provided at least three stories where the participant: 1) witnessed a sanction, 2) sanctioned others, and 3) were sanctioned by others for violating a social norm on an SNS.

4 FINDINGS: NORMS OF (IN)APPROPRIATE CONTENT AND BEHAVIORS ON SNS

Before describing *how* students sanction violations of social norms on SNS, we first detail *what* these implicit social norms are and how they differ across SNS platforms. Some of the content

and behaviors that participants identified as inappropriate (for which they sanctioned others or were sanctioned by others) were consistent with norm violations reported in previous studies [40, 64]. However, while “breaching others’ privacy” is generally pointed out as a norm violation on Facebook [40, 64], these studies do not elaborate on the ways in which violations occurred. When our participants were asked to identify inappropriate content and behaviors on the different SNS that they use, participants disapproved of sharing information that could compromise others’ privacy. This includes sharing content without consent (e.g., sharing screenshots from Snapchat or other re-sharing from one medium to another), releasing personal information about others, harming someone’s or one’s own image/reputation (e.g., sharing embarrassing or risqué photos; strongly worded, opinionated posts; and online personal attacks), and tagging someone inappropriately (e.g., in provocative pictures or revealing their location).

Maintaining one’s privacy, and being considerate of the privacy of one’s friends and family members in what individuals share about themselves or others, was an essential goal for most participants [54, 64]. A small portion of participants (N=4) were also concerned about people they did not know.

In addition to concerns about direct violations of privacy norms, we elaborate on two themes that were salient in our interviews: (1) inappropriate privacy ‘intrusions’ – interruptions to “one’s activities through the unwanted presence or activities of another person” [86, p. 553] – upon our participants’ online space and (2) participants’ perceptions of social norms across different SNS they use.

4.1 It Intrudes on My Personal Online Space

While participants believed that others had the right to post whatever they wanted, they also thought some posts should not be shared online. Participants identified various types of content as inappropriate such as content of a sexual (e.g., nude pictures and people wearing revealing clothes) or illegal nature. They also mentioned that “too much” daily content is inappropriate across all SNS, especially on Facebook. Regarding Facebook and Twitter, participants also stated that negative posts, strongly-worded or opinionated posts, and posts attacking others were inappropriate. Others’ inappropriately shared content did not actually impact participants’ *informational* privacy. However, this content did invade their privacy, as participants argued that it ‘intruded’ on their online personal space (e.g., newsfeeds, timelines, and subscriptions), exemplifying Solove’s [86, p. 553] explanation that such violations are “invasions or incursions into one’s life. It disturbs the victims’ daily activities, alters their routines, destroys their solitude, and often makes them feel uncomfortable and uneasy.” For example, P3 said: “*I didn’t really talk to her as much, and just the fact that I disagreed so strongly with what she was trying to say, I figured I didn’t really need to see that on my feed all the time.*” When participants’ navigation on the site was intruded upon by others’ content, participants were likely to engage in sanctioning behaviors. Thus, participants do believe that others have the right to post whatever they want – as long as it doesn’t violate the viewers’ privacy by violating their expectations of appropriateness. Our participants expected compliance with norms of sharing on different SNS.

4.2 Violations and Contexts: Perceptions of social norms across SNS platforms

We asked participants how they shared content differently across the different SNS. We also asked them to describe a content-sharing incident that was inappropriate in specific social media. Participants justified their sharing decisions or the inappropriateness of a specific sharing incident by describing the *norms of use* applied on different SNS. Participants discussed four factors that determined what information they wanted to share on a particular platform: (1) degree of publicness; (2) expected audiences; (3) reasons for use; and (4) the technical tools available for managing who

Sanction	Online	Invisible	Collective
Confronting	No	No	Yes/No
Gossiping and complaining	No	Yes	Yes
Passively reacting by ignoring/letting it go	No	Yes	No
Removing the violators/violated content	Yes	Yes/No	Yes/No
Implying disagreement by not actively approving	Yes	Yes	No
Uncoordinated and coordinated attacks using comments	Yes	No	Yes/No

Table 1. Sanctions for managing content sharing and privacy in SNS

can see their posts. Our participants saw Facebook as the most public platform and Snapchat as the most private platform, whereas Twitter and Instagram fell in-between these two.

Therefore, although participants identified similar types of inappropriate content sharing or behaviors across different SNS, sometimes (N=4) the *severity* of violating norms of sharing depended on *where* it occurred. This can even, sometimes, impact the strategy to sanction the inappropriate content sharing: “Someone posted a video of me [on Snapchat] when I was drunk, but I wasn’t as upset about it as [I would have been] if it would be on Facebook because it disappears.” (P5). While such a photo of P5 might still tarnish her image, sharing it on Snapchat – where the content is ephemeral and more private because of the limited audience – made her more willing to tolerate and ignore such sharing.

Having covered the social norms of inappropriateness on SNS, in the following section we will explore how people sanction each other when they violate the norms of proper sharing and behaviors on social networking sites.

5 FINDINGS: SOCIAL SANCTIONING STRATEGIES

When participants felt that their expectations for appropriate posting were violated, they tended to apply sanctions, showing that specific content was not welcome in a specific online space. Participants relied on several strategies to manage violations. Building on categories from prior research [7, 23, 54, 54, 78, 102] on the regulation of online boundaries and privacy management, we classified these strategies into the following categories: (1) on-site and off-site sanctions; (2) individual and collaborative sanctions. Based on our findings, we also contribute the new category of (3) visible and invisible sanctions. These three dimensions represent (1) *where* the sanction takes place – on-site, where the violation occurred or off-site, away from the online space in which the violation occurred; (2) *who* applies the sanction – a person or a group; and (3) *how* apparent the sanction is to the violator – visible or invisible. Each sanctioning strategy can exhibit multiple characteristics – for example, “blocking” prevents an individual from seeing, posting, and replying to comments on one’s account. It is a *visible, on-site* sanction that could be applied *individually* or *collaboratively*. Table 1 illustrates the list of sanctions used by our participants and their categories. We elaborate on the content of these categories and their subcategories below. Because our main interest is in impacting the technological affordances in SNS that allow viewers to sanction inappropriate content and behaviors, we begin by organizing our findings under the first dimension: off-site and on-site sanctions.

5.1 Off-site Sanctions

Participants depend on a range of strategies to sanction inappropriate behaviors. Some of these sanctions took place *off-site* away from the social media platform where the inappropriate behavior

originally occurred. These types of sanctions happen either face-to-face or via another communication medium (e.g., phone calls and text messaging). This ‘modality switching’ – the “shifting of interactions from one communication channel to another” [77, p. 288] – is part of everyday boundary regulation strategies. Due to the lack of collaborative privacy controls on SNS, this shift from on-line to off-site channels was a ‘workaround’ that participants used in order to sanction inappropriate online behavior [78] (e.g., requesting the deletion of a photo). Participants also tended to use off-site channels when they believed that the current SNS functionality (e.g., blocking and unfriending) might do more harm than good (e.g., by affecting the quality of one’s relationships with others in an undesirable way).

We found that our participants depended on three general categories of off-site sanctions: confronting the violator, gossiping and complaining about the violator/violation, and reacting only passively to the violator and their unacceptable post.

5.1.1 Confronting. When the shared content might compromise someone’s privacy, individuals prefer to negotiate by confronting the violator [78]. Participants *directly* confront the violator face-to-face or by using private messages. They do so to restore their own privacy or admonish the violator for the violator’s own sake, e.g., when the shared content could possibly affect the violator’s image or reputation. While directly confronting violators, participants avoided doing so ‘publicly’ – online or in-front of other people. As one participant put it, “*I wouldn’t want to get in a fight on Facebook that is then written down and in public*” (P5). The goals of confronting violators vary, from restoring one’s own privacy to actually looking out for the violator.

Restoring one’s own privacy. The majority of our participants had directly confronted and negotiated with a friend or a family member to restore their sense of privacy. Likewise, they reported having been confronted when they violated others’ privacy. For the most part, participants gave their friends “*the benefit of doubt*” (P6), meaning they believed friends did not intend to violate their privacy or harm their reputation. Nonetheless, they explicitly confronted friends to set expectations about their behavior and/or to ask them to delete the inappropriate post. P18 texted her friend to delete an unflattering photo of her on Instagram:

[On] my birthday someone posted on Instagram an old picture of me in like seventh grade from my Facebook, and I just looked so bad. It was like I was making a really weird face and I had braces, and I was not just embarrassed. It was really bad. I think I texted her and said, “That’s so bad. Can you please take it down?” (P18)

P18 did not want to affect her image by the “bad” photo that would stay on Instagram forever.

Admonishing violators for the violator’s own sake. While participants in this category still confronted the violators off-site to sanction their inappropriate behavior, here, the *intention* was instead to protect their friends’ (i.e., the violators’) reputation from their inappropriate behavior. P12 describes below approaching her friend:

[It] happened where someone posted a photo with a caption that didn’t correspond with the photo they posted, and they clearly posted a naked photo on accident. That seemed terrifying and so embarrassing and I texted them. I was like, “Okay, there’s a photo. I think you may have accidentally posted something instead of a different photo.” It was an accident, and it was taken down immediately. (P12)

P12 was aware of a photo her friend shared that could possibly put her in an embarrassing situation with her online friends and followers.

Participants felt that they had a duty to warn friends about the possible negative consequences of their inappropriate posts. Nonetheless, they confronted others *carefully* and *politely*; maybe because participants realized that friends might not always appreciate their advice. Some participants

described their friends' reaction to their advice as terse, e.g., 'Noted. Thank you.' Their friends did not get angry because of their intervention, but they also did not necessarily consider such advice.

In a few cases, participants even *collaborated* with other friends to advise a mutual friend about the possible negative consequences of posts:

[S]he has posted things that were just straight out illegal to do, and she's put them on Snapchat . . . , but some of my friends we had to talk to her and be like "You can't be doing that or posting it because it's very inappropriate." She was more confident in her side, saying that it's her life, and she can share what she wants, which is true. We were just looking out for her. (P3)

Even with this collaborative effort, P3's friend evidently did not appreciate the advice.

Despite the intention of confronting the violator, this *visible* sanction, applied *individually* or *collaboratively*, requires mutual attention and willingness to take into account others' concerns over privacy. Participants did not only care about their own privacy and image, but they were also considerate of their friends' image and made entreaties to them when necessary. It is not surprising that this modality switching and the use of more than one medium in communicating was more prominent among those with strong ties versus weak ties [38].

5.1.2 Gossiping and Complaining. Sometimes participants wanted to avoid an online argument with the norm violators, or they felt helpless to do anything about what had been posted. Therefore, the majority of our participants (N=20) chose to *invisibly* sanction the violator by getting in contact with their friends and relatives off-site, away from the SNS platform, to discuss what happened. Participants gossiped, in other words, with people who would sympathize or share their point of view about posts with which they disagreed. As one participant put it, "*some things do rub people the wrong way, and I understand that they just have to vent about it to feel better*" (P3). In fact, when P3 disagreed with her friend's opinionated post about people's sexual orientation, she said: "*I didn't post anything back. I had just been talking to my roommate about it because we're on the same page*" (P3). Instead of leaving any online comments, she chose to gossip to another like-minded friend.

Many times, participants made screenshots to serve as evidence when they witnessed inappropriate behaviors, using them to "*reinforce the fact that it was unacceptable*" (P10). Participants also wanted to record others' violations for affirmation to, "*make sure that, is it [just] me who is feeling this, or are others feeling it as well, feeling the same thing that I'm feeling?*" (P10). Technological tools (e.g., storing chat logs and screenshots) were used to increase the degree of social control possessed by gossiping witnesses [31]. P2 took a screenshot of her casual friend's Snapchat post to show her friends because she thought the pictures were "*too extreme*," even for Snapchat: "*I showed and told a few friends about this [inappropriate pictures], and we were just like 'Wow she should not have been posting this on Snapchat'*" (P2).

Some participants complained to others after they felt uncomfortable or even angry because their privacy had been violated, but they could not do anything about it since violators insisted that they had the right to decide what to do with their disputed content. Participants complained to their friends to avoid conflict with the violator or to avoid confrontation in the first place. When P5's friend violated her privacy by refusing to remove a photo of her drinking on Facebook, she felt helpless and decided to complain to her friends for comfort and affirmation: "*I talked to my other friends about it, and they were upset too, but it's her account, so none of us could take it down.*" (P5)

Gossiping and complaining, an *invisible, collaborative* sanction, thus allows participants to loudly express their opinion and disagreement while seeking affirmation and sympathy from other people, often hiding their dissatisfaction from the violators, themselves.

5.1.3 Passively Reacting by Ignoring/Letting It Go. Sometimes when participants noticed an inappropriate post shared by close friends or family members, they were hesitant to perform any proactive interactions – whether this was confronting the violator or discussing violations with people in their circles. Instead, participants chose to *passively* and *invisibly* sanction the inappropriate behavior by ignoring the violation. Many of our participants (N=13) described “ignoring,” “disregarding,” or “letting it go” as a way to sanction inappropriate behavior in *their own minds* without having to make a bigger deal out of it. In addition to close friends and family members, this sanction was carried out to avoid confrontation with people they did not know well or whom they knew just casually. P4 described this ignoring sanction: “*Unless it’s a close friend, usually I don’t say anything. Usually I’ll ignore it, and I’ll just think to myself, ‘Wow they really probably shouldn’t have posted that.’*” Her strategy is to ignore inappropriate behavior unless the person in question is close enough to feel comfortable confronting.

In contrast, P16’s and P17’s strategy of ignoring involves deliberately switching from the current social media platform where the violation occurred to another medium or place. P16 explained how this strategy is easier than taking an active position toward this violation:

In general, probably switch social media. So if I’m on Instagram and someone is posting something and I’m tired of seeing the same thing, I’ll just go to Facebook instead or something, or go to Snapchat. . . . It’s easier than causing confrontation and addressing it unless it’s something really bad, then I would take action. (P16)

This strategy embodies an “out of sight, out of mind” philosophy, helping to keep individuals from crossing the line into thinking they should do something more visible about an inappropriate post.

Even when participants wanted to comment in these situations, they hesitated for fear that they might create a heated argument. They preferred to ignore the violation and remained silent: “*I just felt like people that did comment . . . it was never like a civilized discussion. It always just like blew up and created like a bunch of drama, so I wasn’t about to do that*” (P3).

Participants who sanction by *invisibly* ignoring decided to take a minimal reaction toward the behavior or content that they found bothersome or inappropriate in some way. Most of them justified their choice by severity of the violation and how offensive it was; with minor violations (e.g., too many uninterested/daily posts), they found that it was not worth the effort and possible consequences to make an issue out of it.

5.2 On-Site Sanctions

Social networking sites usually provide functionality to their users to control and manage their privacy and boundaries with others. We found that participants used several of these tools (e.g., blocking, unfriending, and online comments), and refrained from using others (e.g., by deliberately not liking a post) in the sanctioning process. However, because most of the on-site sanctions are presumably *visible* to everyone (i.e., to the violator as well as other followers), we found that, in general, participants use these strategies with caution when the person they want to sanction is a close friend or a family member.

5.2.1 Removing the Violators/Violated Content. The majority of our participants have, at least once, chosen to remove the violators or the violated content by blocking, unfriending, or deleting the violator or the content from their social media for various reasons: (1) to end an aggressive online argument, (2) because the friend shared “too much,” (3) because participants disagreed with some of their original or shared posts, or (4) because the shared content harmed their reputation. Blocking, unfriending, and deleting content are presumably *visible* – some more visible than others – to the violator. However, when participants used these features neither the platform nor the sanctioners notified the violators. Therefore, unless the violators perceived certain system cues that such an

action took a place (e.g., the Add Friend button is reactivated) [60], this sanction might remain *unobserved* to the violators.

Blocking, Unfollowing and Unfriending. Participants rarely blocked or unfriended someone with whom they have a strong tie offline. They did not want to affect their relationships with people they are close to, even when struggling with their norm violations:

I think if it was definitely one of my close friends who posted something like that [how same-sex marriage is 'really wrong'], it would be more of an internal struggle to unfriend them because I know they would be upset. (P4)

Instead, participants more readily unfriended or blocked weaker ties – acquaintances, friends of friends, or people they met only online. P3 explained why she decided to unfriend a Facebook friend who posted something with which she disagreed:

[S]omeone from my high school had posted a very strongly worded message about the presidential election and they were trying to convey views that I disagree with, and they were using instances that I thought were taken out of context, so I had unfriended them after I had read it. . . . I hadn't kept in touch with her and I wasn't as close . . . just the fact that I disagreed so strongly with what she was trying to say I figured I didn't really need to see that on my feed all the time. (P3)

P3 wanted to have more control over her personal online space; she did not want content that she disagreed with to appear on her newsfeed and interrupt her navigating experience. Therefore, she decided to unfriend the violator to manage her personal space; because “*there's no way to 'not-like' something on Facebook*” (P3).

Although participants used the terms “delete,” “unfollow,” and “unfriend” interchangeably to mean ceasing to follow someone on social media, in a few cases (N=5), participants used the term ‘unfollow’ to refer to instances when they would stop receiving any posts on their newsfeed from a friend while retaining the person as an online friend. Unfollowing as opposed to unfriending allowed participants to *invisibly* sanction violators for their behaviors. P14 decided to unfollow her grandparents because she did not like their posts: “*On Facebook, I've muted my grandparents because I don't want to see it. . . . They just post the stuff they find amusing and I don't find amusing*” (P14).

P4 explained how unfollowing, as opposed to unfriending, might save people from unnecessary stress in their relationships with close friends:

I think now also that there's an option on Twitter where you can mute people, and so it still says that you follow them because, as stupid as it sounds, if you unfollow one of your friends, they're going to be upset. So sometimes I'll do that [mute people] if they post stuff that I don't like. (P4)

As P4 mentioned, this convenient sanction allowed participants to stop annoying updates on their newsfeed from people with whom they have close ties but do not want to confront or risk adversely affecting their relationship by unfriending them.

Although blocking and unfriending are all primarily *individual* sanctions, some participants (N=5) encouraged their friends or relatives to apply these sanctions on others; some were themselves part of this *collaborative* sanctioning. P14 explained how she kept following and unfollowing a friend who was involved in an online fight with another friend to respect her group decision to do so:

The same girl who was in a fight with my friend, I know I've unfollowed and re-followed her probably six times. I know most of them were for group reasons, but also she's a drama queen. . . . Making it a point to kick that person out of our life sort of thing. It's kind of like a big, “We don't care about you anymore,” which is terrible when you think about it. (P14)

P14's group-level sanction conveyed in this visible way that the friend's behavior was unacceptable, and she was not welcome in their group anymore. Participants usually reported engaging in this kind of practice without warning the violators or letting them know why they were sanctioned.

Deleting Content. Participants also sanctioned inappropriate behavior by deleting the inappropriate comments. Deleting one’s own content in order to simultaneously delete others’ reactions to it is another way to express strong disagreement with violators’ inappropriate behaviors:

A friend of a friend, he was running for . . . some type of congress position but it was more of like a student type thing. And one of his friends went really far back into his Facebook and commented on something really dumb. [I]t was just something that he wouldn’t want to bring up to the surface now that he was in the spotlight and they commented on it. And so it went back on everyone’s timeline and he was really angry about it. . . . [H]e ended up deleting the entire post (P4)

Because the friend’s recent comment made the old post re-appear in the other friends’ timelines, and in order to maintain self-presentation for the expected job position P4’s friend decided to delete the entire post.

5.2.2 Implying Disagreement by Not Actively Approving. Sometimes participants sanctioned violations by intentionally not approving the inappropriate content. A few participants (N=5) described how, just as some online could imply affirmation and agreement (e.g., liking and retweeting/resharing) with inappropriate behavior, the absence of these signals could be used to imply disapproval and disagreement. P18 described a situation when her friend posted a half-naked picture of his girlfriend without her knowledge, stating that liking that picture was as bad as posting it: *“There were people that were liking it, and I thought that was kind of similar to the person posting. If you’re liking it, you’re supporting it, and I didn’t think that was appropriate” (P18)*. P14’s friend retweeted a tweet about drugs that she likes, and P14 mentioned how resharing content conveys an implicit agreement with the content: *“The way I view re-tweeting, and re-blogging, and re-posting is that’s your views that someone else has said, so you’re going to support them by re-tweeting it” (P14)*.

On the other hand, participants described incidents when they intentionally decided to avoid any active reactions (i.e., liking and commenting) to imply disagreement about the violation. P7 explained how she sanctioned her friend by intentionally not replying to her private snaps:

I feel like, on Snapchat, if you’re cool with it, you snap back a response like, “Oh awesome,” or you snap back something similar. So by my not replying to her stories, it was kind of like, “Yeah, I saw them, but I don’t want to engage in that.” . . . Because I feel like, on a private snap, it’s the same as a text message. (P7)

P7, who believes that private snaps are similar to individual text messages – where there is an ‘obligation to reciprocate’ among college students’ SMS usage [95] – decided to ignore her friend’s private snaps to imply that she did not like the content of those snaps. Similarly, when P21’s uncle left a negative comment on his cousin’s political post, he decided to reply to his cousin and intentionally ignored the uncle’s comment: *“I had commented something showing solidarity for that person and being like, ‘I agree with what you say,’ not really talking about what our uncle had commented” (P21)*. P21’s reply to his cousin was a deliberate sanction of his uncle; it conveyed that he did not support the uncle’s point of view.

Similar to passively reacting by ignoring the inappropriate content – deliberately having no reaction toward the norm violations – this sanction might remain totally *invisible* from the violators.

5.2.3 Uncoordinated and Coordinated Attacks Using Comments. Giving direct feedback on inappropriate posts – commenting – is the easiest way to *visibly* sanction people on social media. Despite its ease, participants used commenting with care for people with whom they had strong ties. When participants felt they had to comment to sanction friends and family members, they found ways to indirectly convey their message (e.g., joking, watching their tone, and being polite) to avoid disputes with the violators:

I follow my mom, and from time to time, she makes posts. . . . [S]he posts selfies of her, and I'm just like, "Are you kidding? Selfies are supposed to be for me." I know [that] was mean on social media, but she had made a bunch of selfies, and I was like, "Who do you think you are?" And then all her friends and family members mass jumped on me. . . . And I was just like, "I'm just joking my mom knows I'm joking." Yeah, but I wasn't joking at the same time. (P1)

P1, who left an intentionally rough comment on her mom's selfies, claimed online that she was "joking," even though she was not. A few participants, like P9, explained how their toned down response allowed their friends to think about how others might perceive their post and resulted in their friend deleting the disputed post.

On the other hand, with acquaintances, participants were more willing to give more aggressive, direct feedback:

I shared an article. . . . It was like debunking a lot of myths about abortion as far as development and when life begins. . . . I had an aunt that was like, "Oh, I didn't know you felt this way. Would love to talk with you about it." I respected that because that wasn't like, "Oh, you're wrong." She was very civil. . . . But I did have a classmate (who is very conservative and disagrees) just post a very negative, "I can't believe that you would share this. There's no factual evidence." (P12)

While P12's aunt was polite and indirectly expressed her disagreement about the post, on the contrary, her classmate's response was more direct and aggressive.

While giving feedback is an *individual* strategy people use to convey their disagreement, sometimes it may look like uncoordinated collaboration. Often, when a controversial or inappropriate post was shared, participants expected that the post would be critiqued by audiences. This *uncoordinated collaboration* by others – where people seemed to be ganging up on the person without any explicit collaboration – usually led to a heated online argument. This argument was not only directed at the original poster, who is usually the one that is blamed for posting in the first place [94], but it was also directed toward the commenters. P1 described what happened when a strong comment was left on a culturally offensive post shared by a friend from high school:

I asked [him] to take it down, and I did it in the comments, which is something I usually don't do. . . . [They] weren't even talking about my culture or me as a person, but it was really offensive. A lot of people were in the comments disagreeing with him, and then disagreeing with each other, and it became way bigger than what it needed to be. (P19)

One the other hand, sometimes participants (N=11) felt the need to step up and take an action toward inappropriate behavior; however, simultaneously and, to be more effective, they *coordinated* with others (e.g., with family members and friends) to bring more attention to the violation – mainly by texting and confronting them off-site. Most of the time, these sanctions were designed to protect a person who did not know about a violation and because an individual *alone* could not do anything to stop it. In these situations, participants tried to coordinate with others to sanction the violator (e.g., report and/or attack the violator using online comments). In the following incident, although P18 did not personally take action, her friends kept P18 and others updated as they requested a violator to delete a disputed photo:

[A] guy that went to my high school posted a picture of him in a bathroom with his girlfriend and her back. . . . [S]he was naked, so her whole backside was revealed in the picture. . . . I didn't know if she was aware and so it was even worse thinking that she might not have been aware. . . . [I]t was one of my friends [who] texted me about it. . . . A lot of that was just like people talking outside of Twitter but telling people to go look at Twitter. . . . [S]omeone else texted me that he deleted it. . . . I remember some of my girlfriends were responding publicly to the tweet, saying like, "You should take this down." (P18)

When P18’s friends thought this person violated someone else’s privacy, they stepped up to sanction this person, eventually successfully pressuring him to delete the image. Most of the time, participants reported that this aggressive, publicly coordinated sanction successfully pressured violators to comply with social norms.

6 DISCUSSION

We have found that college students use systematic strategies to sanction norm violations regarding inappropriate content sharing and behaviors. Taking the view that sanctions enforce and sustain social norms, we believe it instructive to first discuss the *effectiveness* of these strategies to sanction norm violations on SNS. We then discuss how design can help to make social norms more *visible* to SNS users. Recognizing that communities are at least partly defined by their expectations for civility and confrontation, for instance, we do not advocate for any particular set of expectations or for the policing of those who violate these expectations, per se. Rather, we wish to provide a means to help support communities in creating whatever kinds of interaction spaces they prefer.

6.1 A Preference for Discreet Sanctions

Despite agreeing on implicit social norms, our participants were often reluctant to take an *active* role in sanctioning violators to enforce norms. As our findings show, regardless of who was sanctioned, most strategies used by participants to sanction norm violations were either *invisible* to violators or *visible but without explanation*. This suggests that current sanctioning strategies might not be effective for users who wish to be warned about their violations to better understand their audience and how to be considerate to other members in the community. In fact, violators may see the *absence of reaction* as implicit support for what they have posted [103].

The alternative is to directly confront norm violations. McLaughlin and Vitak [64, p. 14] note that “confrontation decreased the likelihood that the individual would violate the norm again and ensured that the relationship could continue to be a mutually beneficial one.” However, while our participants confronted friends more than acquaintances [64], in general, they still preferred to avoid *direct* confrontation whenever possible because imposing sanctions has a *social cost* to violators as well as to themselves. Confrontation – whether done off-site or on-site – could result in negative repercussions for relationships or escalate to public fights.

In every interaction, *there is a constant tension between the desire to support the community through sanctioning for norm violations and the simultaneous possibility of undermining one’s relationships with others*. This tension became apparent with participants who especially did not want to negatively affect strong-tie relationships. Therefore, participants refrained from providing direct feedback to their friends and family members unless their own privacy was directly violated. Moreover, when participants did confront those with whom they had strong ties – sometimes for the violator’s own sake – they did so very carefully in both face-to-face and public online comments. Minor violations (e.g., too many posts or overly opinionated posts) were often met with sanctions that were invisible to the violator – e.g., passively reacting, unfollowing, or gossiping to others about the norm violations. The majority of participants resorted to gossip, in fact, because it was the one option that allowed them to maintain a semblance of power and control [45] while still avoiding direct confrontation with family and friends. Gossip plays an important role in conveying information about social norms. Stories about norm violations are often more informative than stories about actions that conform to expectations [4]. Yet, the subject of gossip – the norm violator – is rarely in the loop [6]. Thus, in these instances, too, violators may not learn about their violations and may be less likely to change their behavior in the future.

With weaker-tie relationships, confronting violators was also seen as undesirable. Participants, however, felt neither a desire nor a duty to report violations. Instead, when an acquaintance violated

a norm, whether once or repeatedly, individuals decided to simply ignore, block, or unfriend them to avoid any possible conflict. While blocking and unfriending are ostensibly visible sanctions, they do not actively signal to the violators why the action was taken, much less the exact norm violations committed. Individuals, for example, who consistently post profanity-laden content and/or attacks on others may never know that their behaviors led them to be unfriended by others. They may continue violating norms, remaining unaware of the implicit social norms of their community. Public comments sometimes did not help, either. While some participants responded to an offending post in the comments section to directly sanction the violator, most of the time these public comments were viewed as aggressive and led to heated arguments. For our participants – although perhaps not for others – engaging in heated arguments was itself seen as violating a social norm – an uncivil and unproductive way to sanction violations. Overall, such sanctions may hastily ostracize community members, preventing them from learning to become more civil contributors.

In general, many of our participants preferred an indirect approach, and yet as we argued, many of these actions remain invisible to the violator. Therefore, in addition to approaches that facilitate direct sanctioning and which social media platforms already provide (e.g., commenting and blocking), we argue that the design of online sanctioning tools that avoid direct confrontation and use more indirect methods should be further investigated. One way to facilitate indirect sanctioning, for example, is to design tools that allow ‘plausible deniability’ when applying a sanction. Such indirect mechanisms can be an effective approach to encourage risk-averse users to sanction norm violations while maintaining their privacy if they prefer to hide their actions.

6.2 Designing for Signaling Norm Violations

Excessive online regulations that attempt to design for conformity while failing to understand the dynamic nature of norms on social media [17, 24] can create oppressive spaces with little room for changing or emerging of new norms [71]. By first learning a community’s social norms and expectations of behavior, an individual can then know how to best influence a community to change or challenge its norms. For instance, research has shown that people will intentionally violate norms to instigate change in their community [48, 74]. Therefore, regulations that seek to enforce conformity on SNS risk ostracizing certain members of a community. It can, for instance, put pressure on those holding different values to conform to online norms (e.g., by avoiding sharing certain information or stripping them of their multiple online identities) to avoid social embarrassment and stigma [10]. Regulating SNS can also prevent individuals from sharing stigmatized experiences (e.g., depression or sexual abuse) that are not necessarily inappropriate, but might be anti-normative to their communities [2, 81]. As a result, members may leave and/or seek support in other space(s) [81]. Yet, sanctions that leave violators ignorant of their transgressions are not effective instruments for creating a healthy community, either; they do not help individuals learn what, exactly, are the community’s social norms. The challenge is for creative solutions that can support diverse online environments by providing normative stability as well as the capacity for challenge and change within different kinds of communities.

Villatoro et al. [98] provide an important clue for how this balance can be achieved. They describe two enforcement mechanisms to maintain social norms: ‘punishments’ (where the wrongdoer faces some negative cost such as being banned or being unfriended) and ‘sanctions’ (where a violation is signaled to the wrongdoer with, for example, a warning). Villatoro et al. [98] found that sanctions made social norms more salient and spread more quickly in the community than if the norm violations were enforced only by punishment. Given our findings, we argue that the current design of SNS is ill-suited for effectively *signaling norm violations* while maintaining participants’ goals. *Designing for transparency, rather than conformity*, is a more productive approach. Finding ways to *signal* and make norms more *visible* to the users – without necessarily punishing them for violating

the norm or non-conformity – is a key design opportunity to support sanctioning that does not harm extant relationships or cause conflicts (online or offline), while maintaining the public image of all parties.

Below, we identify implications for designs that we believe can help online community members take an active role to signal norms as well as render these norms more transparent to others without confronting violators or harming social relationships.

6.2.1 Feedback: Between direct and indirect. Feedback, whether it is ‘negative,’ ‘positive,’ or ‘norm reinforcement’ [65] (i.e., teaching norms by rewarding high-quality contributions and punishing low-quality ones) has been shown to play a significant role in producing and sustaining high-quality contributions in online communities [51, 65]. As mentioned in Section 5.1.1, when applying sanctions, people sometimes directly communicate with individuals with whom they have strong ties. While most SNS allow private messages, people struggle to convey the violation to the violator without causing embarrassment or ‘face-threat’. Providing a list of pre-written messages to signal different types of violations might be beneficial because it allows people to confront a user directly without an accusatory message that refers to the situation at hand. Of course, the tone of the message could, itself, provide users with a choice of more direct and indirect face-saving options [11]. For example, if the individual was tagged in an unflattering picture and wanted to send a message to the person who posted the picture, the choices might range from directly asking the violator to remove it to just signaling disapproval of the act. Showing disapproval without directly imposing the removal action minimizes the face-threat and the violator might then voluntarily remove the picture.

6.2.2 Anonymity: Indirectly keeping the peace. One of our goals is to support individuals who primarily want to “keep the peace” with norm violators. Toward that end, we also propose designs that allow viewers to indirectly sanction by providing anonymous feedback that signals the occurrence of a violation to the violator while providing them with ‘plausible deniability’ as to whether they imposed the sanction. With anonymity, people are more likely to give criticism [42, 43]. Friends who observe inappropriate content-sharing by another friend should be able to anonymously report this content and include specific information about the violation. Rather than sending this report to the social media authorities or administrators, this report could then be directly sent to help the violator learn about the violation. Anonymous feedback can also facilitate collaborative ways of sanctioning. Receiving multiple anonymous reports – but with a face-saving, private option – can indirectly but strongly signal to the violator that this behavior is inappropriate and violates community members’ expectations. We envision that feedback about violations could also include the number of requests to remove inappropriate content; the reason behind deleted replies left under others’ posts; and tags such as “too personal,” “invades others’ privacy,” and “too combative.”

Anonymity, moreover, might help with heated arguments that participants found unproductive for sanctioning inappropriate behaviors (Section 5.2.3). Distributed social moderation is used by many online communities, such as Reddit [35], Slashdot [51], and Stack Overflow [62], to allow members to moderate others’ comments and contributions through ratings and/or votes. Such moderation allows interfaces to hide or bury low-quality contributions by default which may be preferred by some users [52] and discourage undesirable behaviors [51]. However, distributed moderation on social networks such as Facebook is more direct since violators can know who moderated their comments, and furthermore lacks down-voting mechanisms. Therefore, a *less direct*, anonymous distributed moderation option could address this issue. Allowing individuals to *anonymously* rate others’ comments and the quality of their posts on SNS can: (1) encourage people to be honest in their moderation without affecting their relationship goals; (2) give other users the opportunity to have a better experience navigating through the site without being disturbed;

and (3) decrease – or at least hide – heated arguments. However, as with any other moderation tool [16], this mechanism might have drawbacks – for example, down-voted people are more likely to down-vote others and negative evaluation can exacerbate trolling behavior [21, 22]. On dominant social media platforms with many users, such as Facebook and Twitter, distributed moderation might not be easy. Designs will need, for example, to address who is allowed to moderate such sites, to what extent anonymous moderators can be impartial in their voting, and the effectiveness of using such mechanisms in different SNS in decreasing heated arguments.

Despite the potential upsides of anonymity for more indirect sanctioning, there are of course potential downsides [47]. In some cases, it may be easy for violators to guess who reported the violation. Further research will be needed to find solutions to effectively anonymize sanctioners' identities – for example, rather than sending the report immediately to the violators, a monthly report can be sent to aggregate all reported violations. Researchers also need to keep in mind how anonymity can sometimes encourage deviant behavior [47, 93] (e.g., cyberbullying [5, 61]) or inflict psychological harm by being too candid. Providing sanctioners with a specific list of common norm violations – e.g., in canned messages (Section 6.2.1) – without allowing them to enter free text might help to mitigate these unwanted consequences. Another possibility is to have the anonymized reports sent to human adjudicators who can try to verify the report, including being in direct contact with the report filer before sanctions are sent to the violator.

6.2.3 Making Support Visible to Sanction Privacy Violations. As Goffman stated, we all need “sympathetic others,” people who have similar experiences to us and “who are ready to adopt [our] standpoint in the world and to share with [us] the feeling that [we are] human and ‘essentially’ normal in spite of appearances and in spite of [our] own self-doubt” [37, p. 19]. Bastiaensens et al. [3] explain that bystanders who observed online harassment have preferences on how to support such victims. They avoid face-to-face communication and prefer to communicate with the victims online using private channels over public ones. However, privately offering support neither sanctions the violators nor helps the community recognize resistance to non-normative behaviors [9, 49, 103]. A study of HeartMob [9] – a platform that allows users to share stories of online harassment and receive support – showed that people who experienced different kinds of online harassment suffered from the lack of online public social support. This lack of support included their close friends and family who failed to understand the severity of their experiences and what they were going through – maybe because they do not see online harassment as a ‘real’ problem [9]. Yet, participants who have experienced such online harassment reported that, over time, the lack of community support made them think that these behaviors are becoming normative on social media, sometimes forcing them to abandon social media or seek support in places where people share similar experiences. In the case of more serious privacy violations, a more *direct* public action might be preferred. For example, individuals who want to support a victim whose privacy has been violated can hit a “support” button to assign a special logo next to the post. The goal here is not to directly sanction the violators or call them out, but to show social support to an individual under duress. This also can make the norm more visible and act as a deterrent against repeat violations. Researchers will need to evaluate the effectiveness of using such features – along with the ability to award them anonymously – to support victims of norm violations. The use of a tool like this in conjunction with the ability to award anonymously is worth consideration.

Lastly, designs that facilitate spaces for people to share their online privacy violations – which are not limited to online harassment [9] – can help community members in two ways: first, online spaces, as opposed to offline gossiping, might increase the chance of violators learning about their violations; second, it also might allow individuals to find more safe spaces to spread the word about violations and get the social support they need from each other to cope with violations [2, 9, 81].

However, designs will need to investigate alternative ways to prevent counter-productive actions against the violator (e.g., social sanctioning in the form of retributive online harassment [8]). There is also danger in outright supporting gossiping, which can lead, for example, to the spread of unfounded accusations [83].

6.2.4 Indirect, Automated Tools for Users to Learn Social Norms. One way to design for indirectness in sanctioning would remove the cost of calling out a violator altogether. Participants commonly worried more about these personal costs of correcting violators’ behaviors than the potential benefits to the violator of learning more about the community’s expectations. What if, instead, automated tools could be designed to inform violators about social norms? Computational models have been proven successful in learning the implicit information sharing norms or behavioral norms of different groups (i.e., contexts) in different SNS [17, 24]; as these tools study users’ actual behaviors and practices, individuals *indirectly* contribute to the process of making norms more salient. Criado and Such [24] studied implicit norms of appropriateness and distribution (i.e., contextual integrity) on SNS and presented an Information Assistant Agent that modelled implicit contexts, relationships, and information sharing norms. Their model was able to infer information sharing norms even if a small proportion of the users followed the norms. Chandrasekharan et al. [17] studied implicit social norms on Reddit by studying comments removed by moderators of subreddits and were able to distinguish between norms that are enforced widely, in specific communities, or exclusively by moderators of individual communities.

Capitalizing on these successful automated models for SNS can help users learn the norms of the different groups and mediums they communicate with and avoid inappropriate information exchanges and undesired information dissemination. On Facebook, for example, individuals can have the ability to activate such a tool for different groups they are part of (e.g., friends, family, and work, or SnapChat vs. Facebook) to study the norms of each group separately. On the “work” group if individuals share “too many” posts daily compared with the average daily sharing of other co-workers, they can receive an automated message stating that their rate of posting is an outlier. This can help individuals distinguish the norms of sharing for this specific group and change their behavior to share more wisely. Among college students – where underage drinking posts were seen as inappropriate among college students [78, 103] – those who post photos with an alcohol-related object or text, especially where they can be seen by family or co-workers, can be notified with the following message: “Sharing a photo with alcohol might be inappropriate or can affect your online self-image. You might reconsider sharing this or share it with a more specific audience.”

These automated tools, moreover, can help participants who refrained from actively reacting on-site – by not liking or resharing – to more directly signal disapproval (Section 5.2.2). For example, Twitter allows individuals to access an activity dashboard to view how their tweets resonate with their audience [96]. Individuals can have access to activities such as how many people saw, liked, or retweeted a tweet. By studying these activities and interactions on individuals’ Twitter accounts, designs may infer if people refrained from “liking” or “resharing” a tweet. For example, if the number of people who expanded a tweet to view more details about it is dramatically less than the average “liking” or “resharing” numbers that an individual’s tweet receives, this feedback may signal that the audience did not like the content of a tweet. However, resharing does not always imply agreement. Therefore, researchers will need to carefully study whether using such an approach can correctly identify passive reactions caused by a norm violation.

Our goal is not to restrict people’s freedom online, but to ‘guide’ them about different social norms of their specific groups that they relate to (e.g., work colleagues vs. close friends). In some cases these hints or ‘nudges’ can improve privacy, e.g., when an inappropriate audience was selected [101].

Giving users “hints” and “suggestions” about these norms might help decrease *unintentional* anti-normative behaviors. However, taking into account the complexity of studying norms, there will be challenges in relying solely on machine learning-based tools to guide users. Current scholarship has highlighted that tools based upon machine learning models end up reinforcing biases inherent in human behavioral data [46]. Researchers will need to carefully test the use of such models and individuals’ satisfaction with them and, of course, always provide the option to use these tools selectively, or simply turn them off.

7 LIMITATIONS

Given that our participants came from a higher education institution in a Western context, we urge caution on prematurely generalizing from our findings. Nevertheless, we found our young adult participants remarkably adept and careful in handling privacy concerns via sanctions. Informed by these savvy users of SNS, we believe our design concepts could transfer to all SNS users regardless of age or other demographic category. Our design concepts broaden the range of responses available to all users when they respond to community members’ posts. Additionally, since our designs are defined more by their functionality rather than any specific form, they can also be made appropriate for a range of different social networking platforms.

8 CONCLUSION

In today’s networked world individuals collaborate to regulate what is shared online. This collaboration between both content creators and viewers serves to shape and reinforce norms on appropriate privacy behaviors and content sharing on social networking sites. Sanctions play an essential role in this process to sustain social norms, thereby protecting privacy. Our study explored how young adults sanction norm violations regarding inappropriate content and their privacy behaviors across different social media.

Through our findings, we systematize sanctioning strategies along three dimensions: *who* performs the sanction (a person or a group), *where* they sanction (on-site or off-site), and *how* apparent the sanction is to the violator (visible or invisible). In studying these strategies, we find that young adults are more than aware that imposing sanctions can come at a personal cost for those who want to maintain their relationships and avoid conflict. As a result, they often sought out discreet – indirect and invisible – sanctions despite their potential to fail in making the violator aware of the violation. We posit that for sanctions to achieve their goals of shaping online behaviors, we need tools that make online norms *visible* to the violator and *signal* violations when they occur without unnecessary burden or direct personal conflict. In the process, all parties might be better able to ‘save face’ and the relationships with each other that they value so much.

9 ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under award CNS-1252697. Rashidi is funded by the College of Computers and Information Systems in Umm Al-Qura University, Saudi Arabia.

REFERENCES

- [1] Irwin Altman. 1975. *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding*. Brooks/Cole, Monterey, CA.
- [2] Nazanin Andalibi, Oliver L. Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In *Proceedings of the 2016 SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, 3906–3918. <https://doi.org/10.1145/2858036.2858096>

- [3] Sara Bastiaensens, Heidi Vandebosch, Karolien Poels, Katrien Van Cleemput, Ann DeSmet, and Ilse De Bourdeaudhuij. 2015. ‘Can I afford to help?’ How Affordances of Communication Modalities Guide Bystanders’ Helping Intentions Towards Harassment on Social Network Sites. *Behaviour & Information Technology* 34, 4 (2015), 425–435.
- [4] Roy F Baumeister, Liqing Zhang, and Kathleen D Vohs. 2004. Gossip as Cultural Learning. *Review of general psychology* 8, 2 (2004), 111–121.
- [5] Andrew V Beale and Kimberly R Hall. 2007. Cyberbullying: What School Administrators (and Parents) Can Do. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas* 81, 1 (2007), 8–12.
- [6] Jörg R Bergmann. 1993. *Discreet Indiscretions: The Social Organization of Gossip*. Aldine, Chicago, IL.
- [7] Andrew Besmer and Heather Richter Lipford. 2010. Moving Beyond Untagging: Photo Privacy in a Tagged World. In *Proceedings of the 2010 SIGCHI Conference on Human Factors in Computing Systems (CHI ’10)*. ACM, New York, NY, 1563–1572. <https://doi.org/10.1145/1753326.1753560>
- [8] Lindsay Blackwell, Tianying Chen, Sarita Schoenebeck, and Cliff Lampe. 2018. When Online Harassment is Perceived as Justified. In *Proceedings of the 12th International AAAI Conference on Web and Social Media*. AAAI Press.
- [9] Lindsay Blackwell, Jill Dimond, Sarita Schoenebeck, and Cliff Lampe. 2017. Classification and Its Consequences for Online Harassment: Design Insights from HeartMob. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW ’17)*. ACM, New York, NY, USA, Article 24, 19 pages. <https://doi.org/10.1145/3134659>
- [10] danah boyd. 2002. *Faceted Id/entity: Managing Representation in a Digital World*. Ph.D. Dissertation. Massachusetts Institute of Technology.
- [11] Penelope Brown and Stephen C Levinson. 1987. *Politeness: Some Universals in Language Usage*. Vol. 4. Cambridge University Press, Cambridge, UK.
- [12] Judee K Burgoon. 1978. A Communication Model of Personal Space Violations: Explication And an Initial Test. *Human Communication Research* 4, 2 (1978), 129–142.
- [13] Judee K Burgoon. 1993. Interpersonal Expectations, Expectancy Violations, and Emotional Communication. *Journal of Language and Social Psychology* 12, 1-2 (1993), 30–48.
- [14] Gary Burnett and Laurie Bonnici. 2003. Beyond the FAQ: Explicit and Implicit Norms in Usenet Newsgroups. *Library & Information Science Research* 25, 3 (2003), 333–351.
- [15] Kelly Erin Caine. 2009. *Exploring Everyday Privacy Behaviors and Misclosures*. Ph.D. Dissertation. Georgia Institute of Technology.
- [16] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. #Thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. In *Proceedings of the 2016 ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW ’16)*. ACM, New York, NY, 1201–1213. <https://doi.org/10.1145/2818048.2819963>
- [17] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. 2018. The Internet’s Hidden Rules: An Empirical Study of Reddit Norm Violations at Micro, Meso, and Macro Scales. *Proceedings of ACM Human-Computer Interaction* 2, CSCW, Article 32 (Nov. 2018), 25 pages. <https://doi.org/10.1145/3274301>
- [18] Nadine Chaurand and Markus Brauer. 2008. What Determines Social Control? People’s Reactions to Counternormative Behaviors in Urban Environments. *Journal of Applied Social Psychology* 38, 7 (2008), 1689–1715.
- [19] Peggy Chekroun and Markus Brauer. 2002. The Bystander Effect and Social Control Behavior: The Effect of the Presence of Others on People’s Reactions to Norm Violations. *European Journal of Social Psychology* 32, 6 (2002), 853–867.
- [20] Elizabeth Chell. 2004. Critical Incident Technique. In *Essential Guide to Qualitative Methods in Organizational Research*. SAGE Publications, 45–60.
- [21] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2014. How Community Feedback Shapes User Behavior. In *Proceedings of the 8th International AAAI Conference on Web and Social Media*.
- [22] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. Antisocial Behavior in Online Discussion Communities. In *Proceedings of the 9th International AAAI Conference on Web and Social Media*.
- [23] Hichang Cho and Anna Filippova. 2016. Networked Privacy Management in Facebook: A Mixed-Methods and Multinational Study. In *Proceedings of the 2016 ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW ’16)*. ACM, New York, NY, 503–514. <https://doi.org/10.1145/2818048.2819996>
- [24] Natalia Criado and Jose M Such. 2015. Implicit Contextual Integrity in Online Social Networks. *Information Sciences* 325 (2015), 48–69.
- [25] Michael A. DeVito, Jeremy Birnholtz, and Jeffery T. Hancock. 2017. Platforms, People, and Perception: Using Affordances to Understand Self-Presentation on Social Media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW ’17)*. ACM, New York, NY, 740–754. <https://doi.org/10.1145/2998181.2998192>

- [26] Paul Dourish and Ken Anderson. 2006. Collective Information Practice: Exploring Privacy and Security as Social and Cultural Phenomena. *Proceedings of Human-Computer Interaction* 21, 3 (2006), 319–342.
- [27] Robert C Ellickson. 1991. *Order Without Law*. Harvard University Press, Cambridge, MA.
- [28] Nicole Ellison, Rebecca Heino, and Jennifer Gibbs. 2006. Managing Impressions Online: Self-presentation Processes in the Online Dating Environment. *Journal of Computer-Mediated Communication* 11, 2 (2006), 415–441.
- [29] Nicole B. Ellison, Megan French, Eden Litt, S. Shyam Sundar, and Penny Trieu. 2018. Without a Trace: How Studying Invisible Interactions Can Help Us Understand Social Media. In *Proceedings of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '18)*. ACM, New York, NY, 129–132. <https://doi.org/10.1145/3272973.3274544>
- [30] Ernst Fehr and Urs Fischbacher. 2004. Social Norms and Human Cooperation. *Trends in cognitive sciences* 8, 4 (2004), 185–190.
- [31] Katleen Gabriels and Charlotte JS De Backer. 2016. Virtual Gossip: How Gossip Regulates Moral Life in Virtual Worlds. *Computers in Human Behavior* 63 (2016), 683–693.
- [32] Vaibhav Garg, Sameer Patil, Apu Kapadia, and L. Jean Camp. 2013. Peer-produced Privacy Protection. In *IEEE International Symposium on Technology and Society (ISTAS)*. 147–154. <https://doi.org/10.1109/ISTAS.2013.6613114>
- [33] Ilana Gershon. 2010. *The breakup 2.0: Disconnecting Over New Media*. Cornell University Press, Ithaca, NY.
- [34] Anthony Giddens, Mitchell Duneier, Richard P Appelbaum, and Deborah Carr. 2006. *Essentials of Sociology*. W. W. Norton, New York.
- [35] Eric Gilbert. 2013. Widespread Underprovision on Reddit. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, 803–808. <https://doi.org/10.1145/2441776.2441866>
- [36] Erving Goffman. 1959. *The Presentation of Self in Everyday Life*. Anchor Books, New York.
- [37] Erving Goffman. 2009. *Stigma: Notes on the Management of Spoiled Identity*. Simon and Schuster.
- [38] Caroline Haythornthwaite. 2005. Social networks and Internet connectivity effects. *Information, Community & Society* 8, 2 (2005), 125–147.
- [39] George Caspar Homans. 1950. *The Human Group*. Harcourt, New York.
- [40] Val Hooper and Tarika Kalidas. 2012. Acceptable and Unacceptable Behaviour on Social Networking Sites: A Study of The Behavioural Norms of Youth on Facebook. *Electronic Journal of Information Systems Evaluation* 15, 3 (2012), 259.
- [41] Christine Horne. 2003. The Internal Enforcement of Norms. *European Sociological Review* 19, 4 (2003), 335–343.
- [42] Craig D Howard, Andrew F Barrett, and Theodore W Frick. 2010. Anonymity to Promote Peer Feedback: Pre-Service Teachers' Comments in Asynchronous Computer-Mediated Communication. *Journal of Educational Computing Research* 43, 1 (2010), 89–112.
- [43] Julie Hui, Amos Glenn, Rachel Jue, Elizabeth Gerber, and Steven Dow. 2015. Using Anonymity and Communal Efforts to Improve Quality of Crowdsourced Feedback. In *Proceedings of the 3rd AAAI Conference on Human Computation and Crowdsourcing (HCOMP '15)*. 72–82.
- [44] Haiyan Jia and Heng Xu. 2016. Autonomous and Interdependent: Collaborative Privacy Management on Social Networking Sites. In *Proceedings of the 2016 SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, 4286–4297. <https://doi.org/10.1145/2858036.2858415>
- [45] Rosabeth Moss Kanter. 2008. *Men and Women of The Corporation*. Basic, New York.
- [46] Matthew Kay, Cynthia Matuszek, and Sean A. Munson. 2015. Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. In *Proceedings of the 2015 SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, 3819–3828. <https://doi.org/10.1145/2702123.2702520>
- [47] Rob Kling, Ya-ching Lee, Al Teich, and Mark S Frankel. 1999. Assessing anonymous communication on the internet: Policy deliberations. *The Information Society* 15, 2 (1999), 79–90.
- [48] Robert E Kraut and Paul Resnick. 2012. Building Successful Online Communities: Evidence-based Social Design. MIT Press, Cambridge, MA, Chapter Regulating Behavior in Online Communities, 125–177.
- [49] Robert E Kraut and Paul Resnick. 2012. *Building Successful Online Communities: Evidence-based Social Design*. MIT Press, Cambridge, MA.
- [50] Cliff Lampe, Nicole B. Ellison, and Charles Steinfield. 2008. Changes in Use and Perception of Facebook. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, New York, NY, 721–730. <https://doi.org/10.1145/1460563.1460675>
- [51] Cliff Lampe and Erik Johnston. 2005. Follow the (Slash) Dot: Effects of Feedback on New Members in an Online Community. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work (GROUP '05)*. ACM, New York, NY, 11–20. <https://doi.org/10.1145/1099203.1099206>
- [52] Cliff Lampe, Erik Johnston, and Paul Resnick. 2007. Follow the Reader: Filtering Comments on Slashdot. In *Proceedings of the 2007 SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, 1253–1262. <https://doi.org/10.1145/1240624.1240815>

- [53] Cliff Lampe, Jessica Vitak, Rebecca Gray, and Nicole Ellison. 2012. Perceptions of Facebook’s Value As an Information Source. In *Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems (CHI ’12)*. ACM, New York, NY, 3195–3204. <https://doi.org/10.1145/2207676.2208739>
- [54] Airi Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. 2011. We’re in it Together: Interpersonal Management of Disclosure in Social Network Services. In *Proceedings of the 2011 SIGCHI Conference on Human Factors in Computing Systems (CHI ’11)*. ACM, New York, NY, 3217–3226. <https://doi.org/10.1145/1978942.1979420>
- [55] Airi Lampinen, Sakari Tamminen, and Antti Oulasvirta. 2009. All My People Right Here, Right Now: Management of Group Co-presence on a Social Networking Site. In *Proceedings of the 2009 ACM International Conference on Supporting Group Work (GROUP ’09)*. ACM, New York, NY, 281–290. <https://doi.org/10.1145/1531674.1531717>
- [56] Maria Knight Lapinski and Rajiv N Rimal. 2005. An Explication of Social Norms. *Communication theory* 15, 2 (2005), 127–147.
- [57] Lawrence Lessig. 1999. *Code and Other Laws of Cyberspace*. Basic Books, Inc., New York.
- [58] Aart C Liefbroer and Francesco C Billari. 2010. Bringing Norms Back In: A Theoretical and Empirical Discussion of Their Importance for Understanding Demographic Behaviour. *Population, Space and Place* 16, 4 (2010), 287–305.
- [59] Eden Litt, Erin Spottswood, Jeremy Birnholtz, Jeff T Hancock, Madeline E Smith, and Lindsay Reynolds. 2014. Awkward Encounters of an Other Kind: Collective Self-presentation and Face Threat on Facebook. In *Proceedings of the 2014 ACM conference on Computer Supported Cooperative Work & Social Computing (CSCW ’14)*. ACM, New York, NY, 449–460. <https://doi.org/10.1145/2531602.2531646>
- [60] Marisela Gutierrez Lopez and Saila Ovaska. 2013. A Look at Unsociability on Facebook. In *Proceedings of the 27th International BCS Human Computer Interaction Conference (BCS-HCI ’13)*. British Computer Society, Swinton, UK, UK, Article 13, 10 pages.
- [61] Davan Maharaj. 1997. UCI Internet Hate Mail Case Ruled a Mistrial. <https://www.latimes.com/archives/la-xpm-1997-nov-22-mn-56652-story.html> Accessed Apr 4, 2019.
- [62] Lena Mamykina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. 2011. Design Lessons From the Fastest Q&a Site in the West. In *Proceedings of the 2011 SIGCHI Conference on Human Factors in Computing Systems (CHI ’11)*. ACM, New York, NY, 2857–2866. <https://doi.org/10.1145/1978942.1979366>
- [63] Alice E Marwick and danah boyd. 2014. Networked privacy: How Teenagers Negotiate Context in Social Media. *New media & society* 16, 7 (2014), 1051–1067.
- [64] Caitlin McLaughlin and Jessica Vitak. 2012. Norm Evolution and Violation on Facebook. *New media & society* 14, 2 (2012), 299–315.
- [65] Jae Yun Moon and Lee S Sproull. 2008. The Role of Feedback in Managing the Internet-Based Volunteer Work Force. *Information Systems Research* 19, 4 (2008), 494–515.
- [66] Jonathan T. Morgan and Anna Filippova. 2018. ‘Welcome’ Changes?: Descriptive and Injunctive Norms in a Wikipedia Sub-Community. *Proceedings of the ACM Human-Computer Interaction 2*, CSCW, Article 52 (Nov. 2018), 26 pages. <https://doi.org/10.1145/3274321>
- [67] Alison R. Murphy, Madhu C. Reddy, and Heng Xu. 2014. Privacy Practices in Collaborative Environments: A Study of Emergency Department Staff. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW ’14)*. ACM, New York, NY, 269–282.
- [68] David M Newman. 2018. *Sociology: Exploring the Architecture of Everyday Life*. Sage Publications.
- [69] Christena E Nippert-Eng. 2010. *Islands of Privacy*. University of Chicago Press, Chicago, IL.
- [70] Helen Nissenbaum. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, Redwood City.
- [71] Karl-Dieter Opp. 2001. How Do Norms Emerge? An Outline of a Theory. *Mind & Society* 2, 1 (2001), 101–128.
- [72] Elinor Ostrom. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.
- [73] Leysia Palen and Paul Dourish. 2003. Unpacking “Privacy” for a Networked World. In *Proceedings of the 2003 SIGCHI Conference on Human Factors in Computing Systems (CHI ’03)*. ACM, New York, NY, 129–136. <https://doi.org/10.1145/642611.642635>
- [74] Sandra Petronio. 2002. *Boundaries of Privacy: Dialectics of Disclosure*. State University of New York Press, Albany, NY.
- [75] Richard A Posner and Eric B Rasmusen. 1999. Creating and Enforcing Norms, with Special Reference to Sanctions. *International Review of Law and Economics* 19, 3 (1999), 369–382.
- [76] Tom Postmes, Russell Spears, and Martin Lea. 2000. The Formation of Group Norms in Computer-Mediated Communication. *Human Communication Research* 26, 3 (2000), 341–371.
- [77] Artemio Ramirez Jr and Shuangyue Zhang. 2007. When Online Meets Offline: The Effect of Modality Switching on Relational Communication. *Communication Monographs* 74, 3 (2007), 287–310.
- [78] Yasmeen Rashidi, Tousif Ahmed, Felicia Patel, Emily Fath, Apu Kapadia, Christena Nippert-Eng, and Norman Makoto Su. 2018. “You don’t want to be the next meme”: College Students’ Workarounds to Manage Privacy in the Era of

- Pervasive Photography. In *Proceedings of the 14th Symposium on Usable Privacy and Security (SOUPS '18)* (SOUPS '18). USENIX Association, Baltimore, MD, 143–157.
- [79] Carmen Santamaría-García. 2014. Evaluative Discourse and Politeness in University Students's Communication Through Social Networking Sites. *Evaluation in context* 242 (2014), 387.
- [80] Richard Schaefer and Robert LAMM. 1995. *Sociology*. McGraw-Hill, New York.
- [81] Sarita Yardi Schoenebeck. 2013. The Secret Life of Online Moms: Anonymity and Disinhibition on Youbemom.com. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*.
- [82] Joseph Seering, Robert Kraut, and Laura Dabbish. 2017. Shaping Pro and Anti-Social Behavior on Twitch Through Moderation and Example-Setting. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, 111–125. <https://doi.org/10.1145/2998181.2998277>
- [83] Linda Shallcross, Michael Sheehan, and Sheryl Ramsay. 2008. Workplace Mobbing: Experiences in The Public Sector. *Workplace Mobbing: Experiences in the public sector* 13, 2 (2008), 56–70.
- [84] Christina Parajon Skinner. 2011. Unprofessional Sides of Social Media and Social Networking: How Current Standards Fall Short. *SCL Rev.* 63 (2011), 241.
- [85] Aaron Smith and Monica Anderson. 2018. Social Media Use in 2018. <https://www.pewinternet.org/2018/03/01/social-media-use-in-2018/> Accessed Mar. 31, 2019.
- [86] Daniel J Solove. 2005. A Taxonomy of Privacy. *U. Pa. L. Rev.* 154 (2005), 477–564.
- [87] Daniel J Solove. 2007. *The Future of Reputation: Gossip, Rumor, and Privacy on the Internet*. Yale University Press, New Haven, CT.
- [88] Anna C Squicciarini, Heng Xu, and Xiaolong Luke Zhang. 2011. CoPE: Enabling Collaborative Privacy Management in Online Social Networks. *Journal of the Association for Information Science and Technology* 62, 3 (2011), 521–534.
- [89] Anselm Strauss and Juliet Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications, Thousand Oaks, CA.
- [90] Fred Stutzman and Woodrow Hartzog. 2012. Boundary Regulation in Social Media. In *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, 769–778. <https://doi.org/10.1145/2145204.2145320>
- [91] Fred Stutzman and Jacob Kramer-Duffield. 2010. Friends Only: Examining a Privacy-enhancing Behavior in Facebook. In *Proceedings of the 2010 SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1553–1562. <https://doi.org/10.1145/1753326.1753559>
- [92] Jose M Such, Joel Porter, Sören Preibusch, and Adam Joinson. 2017. Photo Privacy Conflicts in Social Media: A Large-scale Empirical Study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, 3821–3832.
- [93] John Suler. 2004. The online disinhibition effect. *Cyberpsychology & behavior* 7, 3 (2004), 321–326.
- [94] Lee Taber and Steve Whittaker. 2018. Personality Depends on The Medium: Differences in Self-Perception on Snapchat, Facebook and Offline. In *Proceedings of the 2018 SIGCHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, Article 607, 13 pages. <https://doi.org/10.1145/3173574.3174181>
- [95] Alex S. Taylor and Richard Harper. 2003. The Gift of the Gab?: A Design Oriented Sociology of Young People's Use of Mobiles. *Computer Supported Cooperative Work* 12, 3 (July 2003), 267–296. <https://doi.org/10.1023/A:1025091532662>
- [96] Twitter. 2019. About your activity dashboard. <https://help.twitter.com/en/managing-your-account/using-the-tweet-activity-dashboard> Accessed June. 26, 2019.
- [97] Suvi Uski and Airi Lampinen. 2016. Social Norms and Self-presentation on Social Network Sites: Profile Work in Action. *New media & society* 18, 3 (2016), 447–464.
- [98] Daniel Villatoro, Giulia Andrighetto, Jordi Sabater-Mir, and Rosaria Conte. 2011. Dynamic Sanctioning for Robust and Cost-Efficient Norm Compliance. In *Proceedings of the 22nd International AAAI Conference on Joint Conference on Artificial Intelligence*. AAAI Press, Barcelona, 414–419.
- [99] Joseph B Walther, Brandon Van Der Heide, Lauren M Hamel, and Hillary C Shulman. 2009. Self-generated Versus Other-generated Statements and Impressions in Computer-Mediated Communication: A Test of Warranting Theory Using Facebook. *Communication Research* 36, 2 (2009), 229–253.
- [100] Joseph B Walther, Brandon Van Der Heide, Sang-Yeon Kim, David Westerman, and Stephanie Tom Tong. 2008. The Role of Friends' Appearance and Behavior on Evaluations of Individuals on Facebook: Are We Known by the Company We Keep? *Human Communication Research* 34, 1 (2008), 28–49.
- [101] Yang Wang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, and Lorrie Faith Cranor. 2013. Privacy Nudges for Social Media: An Exploratory Facebook Study. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, New York, NY, 763–770. <https://doi.org/10.1145/2487788.2488038>
- [102] Pamela Wisniewski, Heather Lipford, and David Wilson. 2012. Fighting for My Space: Coping Mechanisms for SNS Boundary Regulation. In *Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, 609–618. <https://doi.org/10.1145/2207676.2207761>

- [103] Loreen Wolfer. 2014. They shouldn’t post that! Student Perception of Inappropriate Posts on Facebook Regarding Alcohol Consumption and The Implications for Peer Socialization. *Journal of Social Sciences* 10, 2 (2014), 77–85.
- [104] Heng Xu. 2011. Reframing privacy 2.0 in online social network. *Constitutional Law* 14 (2011), 1077.
- [105] Nick Yee, Jeremy N Bailenson, Mark Urbanek, Francis Chang, and Dan Merget. 2007. The Unbearable Likeness of Being Digital: The Persistence of Nonverbal Social Norms in Online Virtual Environments. *CyberPsychology & Behavior* 10, 1 (2007), 115–121.

Received October 2019; accepted January 2020