When Forcing Collaboration is the Most Sensible Choice: Desirability of Precautionary and Dissuasive Mechanisms to Manage Multiparty Privacy Conflicts

Mauro Cherubini, Kavous Salehzadeh Niksirat, Marc-Olivier Boldi, Henri Keopraseuth, Jose M. Such, Kévin Huguenin

When Forcing Collaboration is the Most Sensible Choice: Desirability of Precautionary and Dissuasive Mechanisms to Manage Multiparty Privacy Conflicts

MAURO CHERUBINI, University of Lausanne, Switzerland
KAVOUS SALEHZADEH NIKSIRAT, University of Lausanne, Switzerland
MARC-OLIVIER BOLDI, University of Lausanne, Switzerland
HENRI KEOPRASEUTH, University of Lausanne, Switzerland
JOSE M SUCH, King's College London, UK
KÉVIN HUGUENIN, University of Lausanne, Switzerland

Individuals share increasing amounts of personal multimedia data, exposing themselves (uploaders) as well as others (data subjects). Non-consensual sharing of multimedia data that depicts others raises so-called multiparty privacy conflicts (MPCs), which can have severe consequences. To limit the incidence of MPCs, a family of *Precautionary* mechanisms have recently been developed that *force* uploaders to collaborate with the other data subjects to prevent MPCs. However, there is still very little work on understanding *how* users perceive the precautionary mechanisms together with which ones they prefer and why. In addition, precautionary mechanisms have some limitations, e.g., they require linking content to the co-owners' identity. Therefore, we also explore alternatives to precautionary mechanisms and propose a new class of solutions—*Dissuasive* mechanisms—that aim at deterring the uploaders from sharing without consent. We then present a user-centric comparison of precautionary and dissuasive mechanisms, through a large-scale survey ($N = 1792$; a representative sample of adult Internet users). Our results showed that respondents prefer precautionary to dissuasive mechanisms. These enforce collaboration, provide more control to the data subjects, but also they reduce uploaders' uncertainty around what is considered appropriate for sharing. We learned that threatening legal consequences is the most desirable dissuasive mechanism, and that respondents prefer the mechanisms that threaten users with immediate consequences (compared with delayed consequences). Dissuasive mechanisms are in fact well received by frequent sharers and older users, while precautionary mechanisms are preferred by women and younger users. We discuss the implications for design, including considerations about side leakages, consent collection, and censorship.

CCS Concepts: • **Security and privacy → Usability in security and privacy**; • **Human-centered computing**;

Additional Key Words and Phrases: dissuasive strategies; dual system theory; interdependent privacy; multiparty privacy conflicts; online social networks; privacy

Authors' addresses: Mauro Cherubini, mauro.cherubini@unil.ch, University of Lausanne, Faculty of Business and Economics (HEC), 1015 Lausanne, VD, Switzerland; Kavous Salehzadeh Niksirat, kavous.salehzadehniksirat@unil.ch, University of Lausanne, Faculty of Business and Economics (HEC), 1015 Lausanne, VD, Switzerland; Marc-Olivier Boldi, marc-olivier.boldi@unil.ch, University of Lausanne, Faculty of Business and Economics (HEC), 1015 Lausanne, VD, Switzerland; Henri Keopraseuth, henri.keopraseuth@gmail.com, University of Lausanne, Faculty of Business and Economics (HEC), 1015 Lausanne, VD, Switzerland; Jose M Such, jose.such@kcl.ac.uk, King's College London, Department of Informatics, WC2R 2LS London, UK; Kévin Huguenin, kevin.huguenin@unil.ch, University of Lausanne, Faculty of Business and Economics (HEC), 1015 Lausanne, VD, Switzerland.

2573-0142/2021/4-ART53
https://doi.org/10.1145/3449127

ACM Reference Format:
Mauro Cherubini, Kavous Salehzadeh Niksirat, Marc-Olivier Boldi, Henri Keopraseuth, Jose M Such, and Kévin Huguenin. 2021. When Forcing Collaboration is the Most Sensible Choice: Desirability of Precautionary and Dissuasive Mechanisms to Manage Multiparty Privacy Conflicts. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 53 (April 2021), 36 pages. https://doi.org/10.1145/3449127

1 INTRODUCTION

Individuals share ever increasing amounts of personal data online. They usually do so to be visible and relevant to others and to project a positive image of the self through shared data [66, 88, 143]. This is facilitated by the use of smartphones and by the use of dedicated online sharing platforms, including online social networks (OSNs) which accept a variety of multimedia content (e.g., pictures, videos). Unfortunately, sharing such data online may have privacy consequences for individuals, which can, in addition, also lead to severe impacts such as discrimination and cyberbullying [25, 46, 75, 91, 134, 146, 149].

Most research on privacy in OSN focuses on the consequences that individuals face when they expose *themselves* by sharing personal information [35, 46]. Yet, in many cases, individuals also expose others when sharing. Typical scenarios include sharing group photos (*co-owned* data) during social events [76, 138]. Such privacy issues are often referred to as multiparty privacy conflicts (MPCs) and as interdependent privacy situations [28, 60, 137]. MPCs are particularly problematic as the individual who makes the sharing decision (i.e., *uploader*) is not the same person as the individuals who suffer the privacy implications (i.e., *data subjects*).[1] To date, most OSNs do not take into account that most shared content is co-owned and they do not afford collaborative tools to seek and obtain approval for publication from co-owners [77, 137]. While the majority of MPCs might occur when the uploader was acting in good faith, or when pictures were taken in public settings, data subjects can still be hurt or impacted significantly.

Research has tried to solve the problem by proposing solutions that can support the data subjects. These solutions are inspired by spontaneous coping practices that people typically follow when dealing with MPCs [24, 77, 107, 152]. In this article, we will refer to this series of solutions as *Precautionary* mechanisms. Precautionary mechanisms, through computational means, automate collaborative practices and force uploaders to collaborate with data subjects or otherwise limit the shared content. Precautionary mechanisms can be divided into two broad categories: (1) those that *modify the item* to be shared, e.g., by blurring the faces of data subjects that do not explicitly consent to sharing the item [62, 101]; and (2) those that *modify the audience* of the item to avoid undesired recipients, e.g., by filtering those who have access to the item [59, 136]. However, precautionary mechanisms have not been evaluated from a user-centric point of view to understand how users perceive them and whether they prefer one approach over another. Additionally, these mechanisms have limitations too, including: i. the requirement to link content to the co-owners' identity, ii. the reliance on the accuracy of face recognition mechanisms, iii. the incapacity to prevent uploaders from sharing elsewhere[2], and iv. the increased effort required to publish some content.

Given these limitations, we explore a new family of solutions that we name *Dissuasive* mechanisms. Dissuasive mechanisms aim at minimizing the occurrence of MPC by supporting uploaders' decision making about sharing co-owned content. These mechanisms aim to make uploaders reflect on the implications of sharing a given item and raise awareness about the consequences of unilateral decisions [8]. Therefore these mechanisms encourage uploaders to either cease their upload or seek consent—before uploading the content. The design of the dissuasive mechanisms was

[1] More specifically, there might be one or multiple data subjects other than the uploader.
[2] However, it must be noted that if such mechanisms are implemented by the most popular social media platforms, uploaders who switch to different sites to intentionally cause harm will have lower reach on less popular sites.

inspired by the dual-system theory [67, 68, 132, 133]. These approaches create frictions [27] in the interaction flow, which work as mini hurdles, making users think twice and being more considerate before sharing. Dissuasive mechanisms have limitations too. For instance, these might *not* provide an effective prevention against users who *deliberately* create harm to others.[3] Hence, we study user preference across these two types of mechanisms (precautionary and dissuasive).

This study explores the use of dissuasive mechanisms to limit the incidence of MPCs. Furthermore, user preference towards precautionary and dissuasive mechanisms, or the advantages and disadvantages of these methods, are mostly undocumented. Thus, we seek to answer the following research questions:

RQ1 What mechanisms (precautionary or dissuasive) would users prefer to deal with MPCs?
RQ2 What specific solutions, if any, within mechanisms (precautionary or dissuasive) would users prefer to deal with MPCs?
RQ3 Would different types of users prefer different mechanisms?
RQ4 Why would users prefer some mechanisms over others?

Our paper contributes to interdependent privacy design by proposing a new family of solutions for addressing MPCs — dissuasive mechanisms. Different from previous methods, our proposed approach aims to avoid MPC before the content is being shared. This is crucial, as resolving an MPC after it occurs may be impossible, particularly in severe cases. Our paper integrates prior work on precautionary mechanisms and compares them with dissuasive mechanisms, both at the level of individual solution and as a family of solutions. We also curated a large-scale ($N = 1792$), representative sample of adult Internet users allowing us to generalize our findings to the US adult population. Importantly, we studied different individual factors such as demographics, attitudes, and MPC related experiences to understand which factors can influence user preference toward each strategy. Last, we discuss implications for the design of tools supporting OSN users to deal with MPCs.

2 RELATED WORK

We discuss three main streams of related literature: (1) research that has uncovered and quantified the privacy risks from MPCs, (2) research that has uncovered, studied, and categorized some of the strategies that users are forced to use to deal with MPCs in the absence of better technical support, and (3) research proposing novel interfaces and methods to help users manage MPCs.

2.1 Privacy Risks of MPCs

Risks stemming from data uploaded on social networks have extensively been studied in the research community. The risks span from image damage and shaming [75, 149] to cyberstalking and cyberbullying [25], identity theft, and from discrimination [46] to revenge pornography [134, 146]. In a survey targeted at Facebook users, Ilia et al. [62] found that 92% of users have access to photos depicting strangers. Hoyle et al. [55] found that users perceived photos portraying two persons less likely to be private than photos showing only one person, indicating that users are not aware of the privacy risks of co-owned content. Li et al. [78] developed a user-centric taxonomy of sensitive photo contents and showed that respondents are likely to share photos depicting events and other people. In a large-scale user study (1033 participants), Such et al. [138] found that almost all respondents experienced an MPC at least once on social networks; almost 60% of the uploaders did not ask for permission before sharing; 30% of the MPCs were never resolved; and approximately 90% of the conflicts were resolved only after the photo had already been shared, with the two most popular conflict resolution strategies being either removing the photo or doing nothing. The

[3]However, warning uploaders in advance about community standards can waive their ability to claim good faith.

question of user awareness has also been studied: Henne et al. [53] found that 52% of users accidentally became aware of shared content where they are portrayed. Olteanu et al. [101] reported that a staggering 41% of Facebook users do not tag or link their friend's profile in their posts; 31% of them declared asking their friends to remove content that they posted and 17% declared asking Facebook to remove such content; 10% of the participants claimed that they were victims of discrimination; and 4% reported being victims of revenge pornography. Lastly, Chutikulrungsee and Burmeiste [26] reported that more than one fourth of their 460 survey participants requested another user to remove a shared photo.

2.2 Conventional Coping Strategies to Manage MPCs

As a result of insufficient support for MPCs in current OSN infrastructures, users employ manual coping strategies [17, 77, 152, 153], such as trying to anticipate consequences for others [77], seeking approval before making a post [77, 116], applying self-censorship or less frequently engaging with OSNs [24, 153], changing their offline behavior [17, 77], imposing sanctions against privacy violators [117], and negotiating privacy policies with other users [17, 77, 116, 152]. Research has extensively studied these strategies that users use to try to cope with MPCs in current OSN infrastructures.

Building upon seminal works by Altman [7] and those by Palen and Dourish [103] about boundary regulation of privacy, Wisniewski et al. [151, 152] constructed a taxonomy with five types of interpersonal boundaries that OSN users employ to manage their privacy, namely: i. disclosure boundary; ii. relationship boundary; iii. network boundary; iv. territorial boundary; v. interactional boundary. While this study provided a general understanding of user boundary regulation behaviors in OSNs, recent studies have focused more on managing privacy issues about co-owned information. For instance, Lampinen et al. [77] identified two dimensions of techniques OSN users applied: (i) individual vs. collaborative and (ii) preventive vs. corrective; The finding showed that individual coping strategies are better supported by OSN systems than collaborative ones. Moreover, users often tried correcting privacy issues rather than preventing them. These findings later were confirmed by Cho and Filippova's study [24] revealing four main constructs of privacy management strategies: (i) information control; (ii) corrective strategies; (iii) collaborative strategies; (iv) preventive strategies. Finally, Such et al. [138] studied a large number of MPCs —caused by sharing photos on OSNs—people experienced and on how these were dealt with. They identified twenty different strategies to resolve the conflict including apologizing, flattering the affected person, and cropping people out of the photo.

Overall, the solutions currently supported by popular OSNs have been shown to be insufficient to successfully deal with MPCs, even when users perform the coping strategies mentioned above to try to work around MPCs. This is due to the fact that privacy breaches may have already occurred [90, 137] before the strategies are enacted, the delay for removing the content may be unsatisfactory [80], the coping strategies may be impracticable [137], there is an incorrect assumption regarding complete user awareness [2, 101], and the coping strategies do not protect against the service provider itself [60]. Given these limitations, researchers developed more automated mechanisms to help users manage MPCs in OSN, which are discussed next.

2.3 Precautionary Mechanisms: Novel Solutions to Manage MPCs

Earlier studies focused on proposing novel methods to help users manage MPCs. This includes automated mechanisms that either modify the audience of the item (i.e., who can see a photo) or modify the item to be shared (e.g., blurring the faces in a photo) to avoid or resolve an MPC. Regarding audience modification, different approaches were considered for co-owners to agree on who should be able to see an item and to verify that only selected individuals can see the photo. This

included studies based on secret sharing [16], the Clarke-Tax mechanism to encourage uploaders to acknowledge co-owners [130, 131], aggregated voting [23, 118, 141], a new framework to reach a consensus based on co-owners' trust values [6], novel access control models [56–59, 145, 150], computational mechanisms for adaptive audience recommendation [135, 136], and AI-based negotiation techniques that use game theory [115, 139], argumentation theory [36, 38, 72, 94] and human values theory [92–94]. Furthermore, Fogues et al. [36, 37] implemented a recommendation system trained with data collected from a user survey. Keküllüoğlu et al. [69, 70] and Rajtmajer et al. [115] considered multiple interactions between OSN users over time. The potential limitation of the audience modification techniques is the possibility of content redistribution, where even trustworthy friends who have access to one's group can redistribute the content to other groups or OSNs. Regarding item modification, several approaches were considered by Ilia et al. [62] who proposed a fine-grained, collaborative multiparty access-control model for photos published on social networks; it relies on face detection, obfuscation (blurring) and individual manual decisions of the subject regarding who can have access to photos. Aditya et al. [5] developed a mobile app for collecting consent for photos that relies on privacy policies, face recognition and homomorphic encryption. Olteanu et al. [101] used a two-party architecture and cryptography to build a privacy-preserving system for sharing photos with consent from all involved users. Ilia et al. [61] designed a collaborative multiparty access control model that ensures considering the privacy preferences of all co-owners of shared content. Hasan et al. [48] developed a computer vision approach identifying bystanders in photos taken in public places to preserve their privacy. The potential limitations of item modification approaches are the possibility to infer data subject identity based on the features involved in the content (e.g., location, clothes, friends), delays in the content to be shared, requiring additional effort to increase the aesthetics of the photos [49, 50, 79], the possibility of false face detection, and possibly rising additional privacy issues by requiring the identification of all subjects within the content [137].

2.4 Dissuasion versus Nudging

Nudges are "*any aspects of the choice architecture that alter individuals' behavior in a predictable way without forbidding any options or significantly changing their economic incentives*" [140, p. 6]. Nudges have been studied in different fields from health [65] to HCI [22], and in particular privacy [10, 83, 89, 147] (see [4] for a review on privacy nudges). While designing dissuasive mechanisms (Section 3.1), we were inspired by privacy nudges. Nevertheless, it is worth noting that there are two distinctions between the privacy nudges and our proposed dissuasive mechanisms. First, privacy nudges are often related to users' subconscious behavior, where users are not necessarily aware of the intervention received. In other words, privacy nudges guide users' behavior without imposing a particular decision (e.g., using framing and default techniques to bias users' privacy decisions [10, 83]). In contrast, dissuasive mechanisms are more perceptible where they increase user awareness, and allow users to reflect on their behavior (e.g., by explicitly warning users). Second, privacy nudges are commonly considered as soft paternalistic interventions (e.g., by gently reminding users about their behavior [89, p. 784], [148]). Dissuasive mechanisms, in contrast, propose harder interventions by informing users about the tangible consequences of their actions (e.g., with fines if they violate the privacy of others).

2.5 Background Legislation Relevant for MPCs

The non-expert reader might wonder why technical solutions are necessary to deal with MPCs or believe that current law is sufficient to protect Internet users. In this section, we briefly discuss the laws of different countries and detail how, unfortunately, these often fall short of protecting citizens sufficiently.

According to the Universal Declaration of Human Rights, everyone has the right to privacy and has the right to the protection of the law against privacy interferences [112, art. 12]. The right to privacy is emphasized at the regional level in Europe [99, art. 8] and in the US [39, art. 11]. Recently, the European Union introduced a specific regulation for (personal) data protection, namely the General Data Protection Regulation, or GDPR (Reg. EU 2016/679 [111], which replaced Reg. 95/46/EC). In the US, there are a variety of federal and state laws for privacy protection (e.g., California Consumer Privacy Act [98]). Although a recent study discussed the implications for future legislation for privacy threads in intimate relationships [13], none of the existing laws clearly mention the case of specific MPCs. In the U.S., one could argue that for example, photography in public settings is always legitimate [74]. Even in a public setting, if the photo context is outside of the community norms, such an act could violate someone's privacy. It must be noted that pictures taken in a public setting might cause two forms of torts [114]: (a) public disclosure of private facts, and (b) diminishing the public image of a person. Therefore, the fact that the picture was taken in a public setting in itself is not sufficient to state that consent to publish the picture is not required (e.g., `Google Street View` case [34]). Indeed, following the growing number of cases, different states in the U.S. introduced specific legislation against cybercrimes and in particular cyberbullying [52]. Last, different countries such as Canada [110], France [44], Italy [29], Japan [84], and a few states in the US [82] have enacted laws for severe cases of MPCs such as revenge pornography. At the rate multimedia content is produced and shared [71], technical solutions have become indispensable to reducing undesirable effects of MPCs.

3 DESIGN

In this section, we explain our design of two large categories of mechanisms for addressing MPCs: *dissuasive* and *precautionary* Mechanisms. We started designing dissuasive mechanisms through brainstorming sessions while we were reading and discussing literature that informed our design decisions. For the precautionary mechanisms, we mainly relied on the earlier technical solutions (cf. Sec. 2.3) and we only designed how these would be presented to the users. We developed the interventions by following a user-centric development process [64], which comprises formative research and a few iterative cycles of design and testing. After designing each solution, we built interactive prototypes,[4] and we modified the prototypes after a pilot study.

3.1 Dissuasive Mechanisms

3.1.1 Design Rationale. We designed *dissuasive* mechanisms to encourage uploaders to behave in a more considerate and mindful fashion before sharing content online to reflect on the consequences of MPCs. This might result in uploaders either ceasing to share content or seek explicit consent from data subjects before doing so. Dissuasive mechanisms do not provide any control to the data subjects, rather they directly intervene on the uploaders.

3.1.2 Design Inspirations (background). To design the dissuasive mechanisms, we rely on the dual-system theory [67, 68, 132, 133]. This theory posits that humans have two types of thinking processes: *System 1* and *System 2*. System 1 is fast and intuitive, and it requires minimum cognitive effort to handle automated everyday tasks. System 2 is slower but more deliberate, conscious and logical. Handling tasks with System 2 requires greater cognitive effort. While System 1 helps individuals to manage everyday behaviors in an efficient autopilot mode, it is prone to error. Thus, switching from System 1 to System 2 can aid an individual when analytical thinking is required, for example, when one faces challenging tasks.

[4]The storyboards of all the mechanisms are available in the Open Science Framework repository. See https://doi.org/10.17605/OSF.IO/NDBK7, last accessed January 2021.

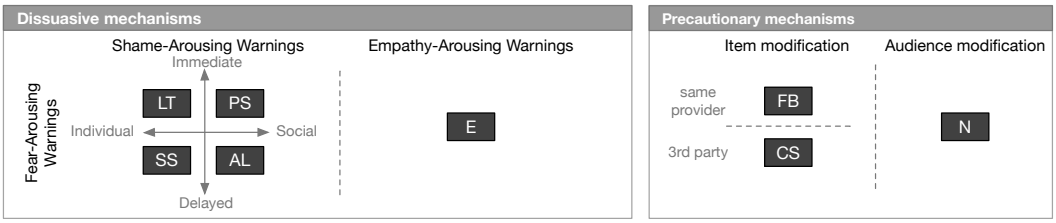


Fig. 1. Contrasting graph illustrating the differences between the mechanisms tested in this study.

Earlier studies [27, 81] argued that users generally use technology mindlessly in a habitual way (using System 1), without thinking about their behavior and its consequences. Although, such fast and automated behavior is useful most of the time (e.g., reacting to a sudden danger while driving), in some situations it can lead to negative outcomes (e.g., responding to a useless mobile phone notification while working on an important essay). Building on the dual-system theory, Cox et al. introduced *design frictions* (or *microboundaries*) in interaction design to support mindful behavior while using technology [27]. Design frictions create tiny pauses, compelling users to switch from System 1 to System 2. In other words, they act as mini hurdles to inhibit users from mindless interactions with technology [81] (e.g., an app that warns users not to check notifications during working hours). While alternatives that focus on System 1 might also be possible, such as alternatives based on *privacy nudges* [4, 10, 83, 89, 147], we focus this work on System 2 interventions following Phelan et al. [108], who showed that, using System 2, users can better describe their privacy concerns and perform better privacy risk assessments. In particular, to avoid mindlessly sharing online, and thus possibly incurring MPCs, we designed frictions in the form of warning interventions. These create a brief moment in which the uploaders might reflect on what they are doing and steer the uploaders away from an automatic mindless sharing and toward more mindful behavior.

The effectiveness of dissuasive mechanisms largely depends on their ability to provoke self-reflection in the mind of uploaders. We followed recent work on *anticipated emotions* (i.e., cognitive predictions about future emotions) and decision making [86]. The basic idea behind this research is that people tend to avoid taking actions that could result in negative emotions (e.g., shame, guilt, sadness) and to pursue those that will result in positive states (e.g., pride, joy). An emerging body of work suggests that the anticipation of emotions could be effectively used in interventions aiming at shaping pro-social behaviors [15, 45, 85, 120, 121, 127, 128, 154]. We therefore designed persuasion mechanisms that anticipate three types of emotions that can provoke self-reflection: (1) *Fear-arousing*, (2) *Shame-arousing*, and (3) *Empathy-arousing* warnings. Given that fear and shame are both negative emotions, we consider these as two orthogonal dimensions and therefore we designed four distinct mechanisms that influenced these two emotions in different manners (cf. Figure 1). The mechanisms we designed might be susceptible to habituation and therefore their effectiveness might decrease in long-term use. We discuss this issue in Sec. 6.3.4.

3.1.3 Fear- and Shame-Arousing Warnings. Decades of psychology research have established that fear appeals (i.e., threats) can persuade individuals to avoid particular actions [120, 121, 154]. Fear appeals have mainly been studied and applied in health-related domains, e.g., warnings such as "*having unprotected sex puts you at risk of acquiring AIDS, a deadly disease*" have been found to encourage safer sexual behavior [120]. Another well-know example is the "*Smoking Kills!*" slogan that aims at raising awareness about the consequences of smoking tobacco. Scholars found that threatening communications with *immediate* consequences (as opposed to *delayed* consequences)

had a greater effect in achieving compliance with the target behavior [1, 51]. Therefore, we hypothesized that warnings that lead to immediate consequences should have higher effectiveness than delayed consequences.[5] Other research has focused on shame appeals. Scholars found that shame is a social emotion that warns the individual that they might be rejected or ostracized in social relationships [41]. Anticipating shame as a direct result of anti-social behavior was successfully used in interventions aiming at supporting pro-environmental decision-making [124]. Sharing multimedia online is known to maintain social presence and improve self-presentation [66, 88, 143]. Therefore, we hypothesized that warnings that threaten social shaming should have higher effectiveness than warnings that threaten consequences that might not become of public knowledge.

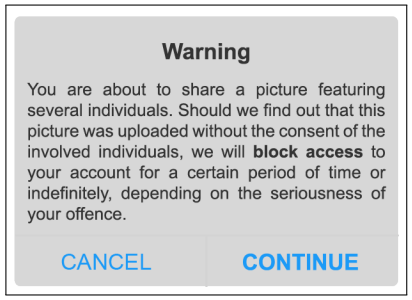
Public Shaming, or PS. This method aims at inducing *social shame* in a *direct* manner. PS threatens uploaders' social reputation by advertising any non-compliant behavior on the news feed of their peers. In short, if an uploader shares multimedia content without the explicit consent of the other individuals involved, a message will be posted on the walls of all his contacts on the social network. The content of the post will indicate that the user misbehaved by sharing photos of their friends without obtaining their consent.

The interaction flow for PS goes as follows: it starts with the uploader sharing a photo through a social network app. Then a pop-up warning message is displayed to seek confirmation from the uploader regarding the sharing intent or to cease the sharing: "*You are about to share a picture featuring several individuals. Should we find out that this picture was uploaded without the consent of the involved individuals, your post will be removed and* all of your contacts will be notified *of your actions. In addition, your profile page will keep track of your offenses.*"

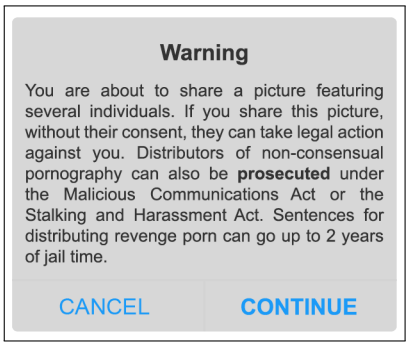
Account Locked, or AL. Similar to PS, AL threatens uploaders' *social life*, but with *delayed* consequences. This method threatens the uploaders to be blocked for a given period of time (e.g., a month), or indefinitely, in the case where a non-consensual sharing was detected/reported. While PS notifies *immediately* the peers of the uploader about the non-consensual sharing, with AL, peers might discover that the uploader was banned only a few days or months *after* the non-consensual sharing has occurred (e.g., when the punished uploader does not respond on the social network where the account was locked). Figure 2(a) presents the warning message.

Law Threat, or LT. This mechanism is designed to threaten uploaders that sharing without consent may lead to an immediate impact to their personal situation, by a monetary fine and/or jail time. In particular, this method reminds the uploader that the other individuals involved may bring a privacy violation to court. This method is grounded in the doctrine of *deterrence* [42, 104] that explains how sanctions may prevent further crimes. Deterrence is a common practice in justice systems [87, p. 20] to reduce different types of offenses through threatening a sanction (cf. [95] for a review on deterrence). The effectiveness of the deterrence strategy highly depends on the perceived certainty and severity of sanctions. If people feel the proposed consequences are uncertain and inconsequential, this approach might fail. A recent study [155] demonstrated the use of the deterrence principle as an effective approach to prevent cyberbullying in social networks. Results showed that when sanctions were more certain, people's intention to engage in cyberbullying was reduced. The warning message used in the experiment, reported in Figure 2(b), referred

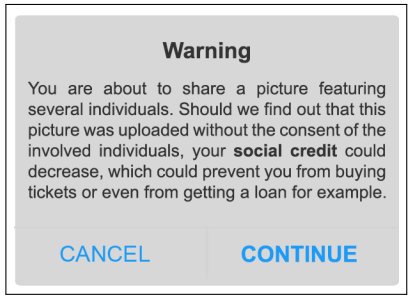
[5]Earlier privacy research studied the effect of the immediacy in feedback exposure for location sharing [105, 106]. Nevertheless, our focus is slightly different as we consider the immediacy in terms of 'consequences' that an uploader might face after sharing a non-consensual content.



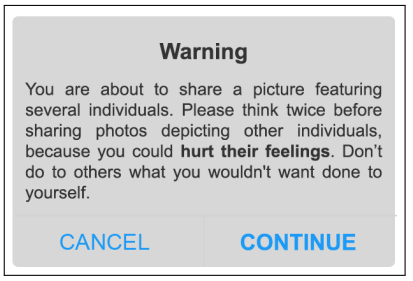
(a) Account Locked Strategy (AL).



(b) Law Threat Strategy (LT).



(c) Social Score Strategy (SS).



(d) Empathy Strategy (E).

Fig. 2. Screenshots showing warnings used in dissuasive mechanisms (AL, LT, SS, E).

to U.K. legislation. The existence of–and the recent trend to pass new–specific laws regarding revenge pornography and cyberbullying in many countries (cf. Sec. 2.5) motivated the inclusion of this mechanism in our study.

Social Score, or SS. Similar to LT, SS is designed to threaten uploaders' personal situation. However, the consequences of sharing without consent will be delayed in time. The mechanism is modeled after the social credit system in China,[6] a state-owned and centralized reputation system where every citizen is provided a rating that decreases if the person commits crimes, and increases if the person performs good deeds. This system was evaluated by a recent survey ($N = 2209$) [73], where Chinese respondents highly approved the strategy. However, it is worth noting that the social credit system remains highly controversial in western societies [18]. In this paper, the Social Score mechanism is that if the uploader is found guilty of sharing content, their social score would decrease, making it more difficult to obtain credit or to use public services (cf. Figure 2(c)).

3.1.4 Empathy-Arousing Warnings. Empathy-arousing communications can be as effective as threatening communications [127, 128]. For instance, empathy has been recently shown to directly influence people to avoid drunk driving [127]. The technique consists in rising a moral dilemma in the mind of the uploaders asking them to think twice before sharing multi-party private information. This technique is supported by the psychological consequences of the *sense of guilt* that individuals might experience if they cause harm to others [45].

[6]See https://en.wikipedia.org/wiki/Social_Credit_System, last accessed January 2021.

Empathy, or E. This method asks the users to put themselves in the shoes of the other person involved. The text reminds the users that, although they might perceive the content as appropriate for sharing, the other individuals depicted might feel otherwise. The warning message makes specific reference to the fact that sharing without consent could *hurt* another individual's feelings and that, in severe cases, it might lead to extreme consequences such as suicide — Figure 2(d).

3.2 Precautionary Mechanisms

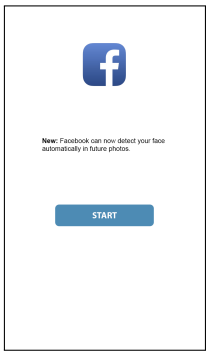
Earlier privacy literature [24, 77, 152] revealed collaborative and preventive strategies as common practices of OSN users to cope with privacy violations (cf. Sec. 2.2). For instance, before sharing a particular content, users negotiate whether the content is appropriate to share and who should have access to it. However, such practices usually occur beyond OSN platforms, face-to-face or through messaging platforms [24]. Lampinen et al. [77] highlighted how we still miss tools that allow OSN users to solve MPC collaboratively. Recently, several studies [59, 62, 101, 136] have proposed precautionary mechanisms; computational mechanisms that automate the collaborative practices (cf. Sec. 2.3).

Precautionary mechanisms allow data subjects to prevent uploaders from sharing without their explicit consent. These mechanisms *force* uploaders to seek consent before the content to be shared is made available to others. Precautionary mechanisms rely on face-recognition algorithms to detect and identify other individuals in the content. Once recognized, these mechanisms match the face of these individuals to their online profile. Finally, they ask the data subjects for consent regarding the publication of the media. Following this interaction scheme, we developed three mechanisms: *FaceBlock* (or FB), *ConsenShare* (or CS), and *Negotiation* (or N). The first two perform an item modification and the last one performs audience modification.

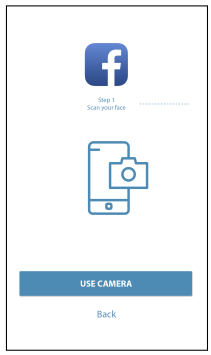
FaceBlock, or FB. This technique builds on earlier studies on item modification [62]. Figure 3 presents the interaction flow of the prototype we designed. [34]At the beginning of the sequence, the user is asked to create his fingerprint by taking a selfie through the camera of the phone. Next, the user is asked to upload a photo of his document (ID, passport, or driver's license) to verify his identity. Then, the user is brought back to the newsfeed page. There he decides to share a photo from the camera roll. The chosen photo is of three individuals (the user and two friends). Upon sharing, the faces of the friends appear blurred as they have not yet given their consent to be shown in the photo being shared.[7] At the same time, the user receives a notification that one of his friends has also posted a photo where he is shown. By tapping on the notification, he can review the photo and decide whether to accept or refuse his face being displayed. As he appears naked, he refuses. Finally, at the end of this sequence, he receives a notification that states one of his friends has accepted to appear on the photo that he previously shared. As a result, the face of the friend is revealed.

ConsenShare, or CS. As we were interested in whether users could see the service provider as an adversary, we contrasted this mechanism with the previous one (FB). This mechanism is modeled after ConsenShare [101], an identity manager for multiple service providers (i.e., OSN such as Facebook or media sharing websites such as PornHub). The main interaction flow of this method resembles the case presented above for FB. However, the first steps of creating a fingerprint of the face and verification of the identity are performed by a service/app that is distinct from the service provider. This ensures that the service provider has no knowledge of the users appearing in photos prior to their granting consent and that the identity manager has no knowledge of the photos being uploaded.

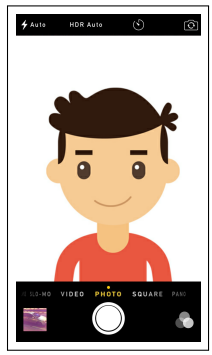
[7]Note that this technique was also used in the recent `Google Street View` case [34].



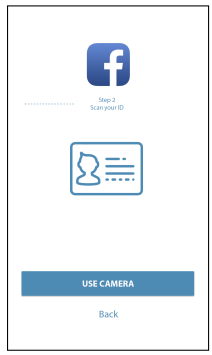
(a) Upon login the user is presented with a promo of the new feature.



(b) The user is asked to take a photo of their face to build the fingerprint.

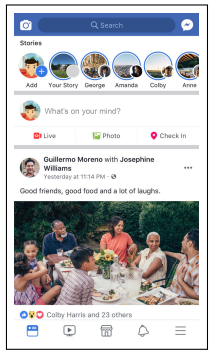


(c) The camera app opens up.

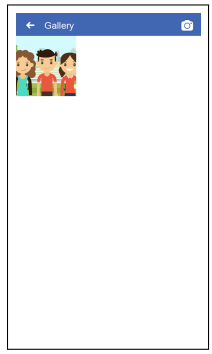


(d) The user is asked to take a photo of the ID to verify their identity.

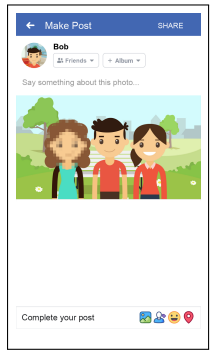
(e) The user takes a photo of the ID.



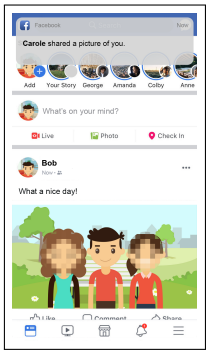
(f) On the main news feed, the user is asked to share a photo.



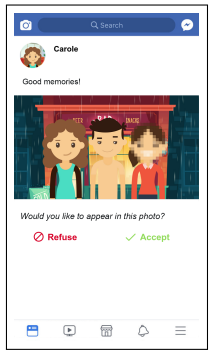
(g) The photo is selected from the camera roll.



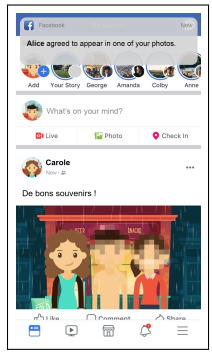
(h) The user sees a preview of the photo and can add a short descriptive text.



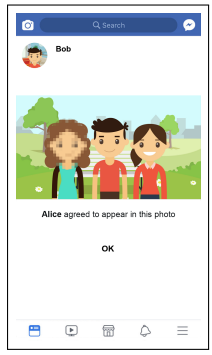
(i) The photo is published on the wall of the user. After a few seconds, a friend of the user shares a different photo and the user receives a notification.



(j) The user checks the photo and judges his appearance inappropriate, therefore he refuses his consent.



(k) The photo appears with his face pixelated. After a few seconds he receives a notification that Alice agrees to appear on the photo that he shared earlier.



(l) The photo now reveals Alice's face, but it is still missing Carole's approval.

Fig. 3. Screenshots showing the interaction flow of the FB prototype.

Negotiation, or N. This technique is modeled after several studies that presented approaches for agreeing on the audience that will apply to multimedia content in OSN, whether by aggregating individual privacy preferences in a predetermined [23, 59, 141] or adaptive way depending on the context and sensitivity of the content [136], or by automatically negotiating the audience [92, 93, 139]. In this paper the N mechanism follows the same basic interaction of the FB, but after the content is flagged as depicting multiple individuals, the solution prompts each member involved to state their preferred sharing settings: 'private', 'only friends in common', 'only friends', and 'public'. Then, to reach a decision, the algorithm combines the individual settings. In this case, it selects the most restrictive privacy settings among all the individual responses as in [141].

4 METHODOLOGY

We conducted an experiment with a two-fold purpose. Our primary goal was to evaluate and compare the dissuasive and precautionary mechanisms introduced above. As a secondary goal, we also collected from participants their MPC experiences and coping strategies in their everyday life. This information helped us to investigate the factors (e.g., demographics, attitudes, MPC related experiences) that influence users' preferences towards each mechanism. We set the following hypotheses:

H1 Users prefer one of the mechanisms (precautionary or dissuasive) to deal with MPC.
H2 Within mechanisms (precautionary or dissuasive), users also prefer one specific mechanism to deal with MPC.
H3 Dissuasive mechanisms that imply immediate consequences are preferred over those that imply delayed consequences.
H4 Specific groups of users based on role, demographics, multimedia sharing practices and MPC experience prefer different mechanisms.

4.1 Experiment Design

In order to achieve the two goals above, we designed a large-scale online survey comprised of two parts. In the first part, we collected data on sharing practices for multi-subject data using a critical incident method [33] (i.e., having respondents remember episodes where they experienced MPCs, as in [138]). In the current study, we tackle all kinds of MPCs; some parts of the study focused specifically on the MPCs that lead to severe consequences (i.e., public shaming and discrimination related to nudity or sexual content). In the second part of the survey, to collect feedback about our mechanisms, we used a *scenario testing* approach with storyboards; a well-established methodology to elicit feedback about technology [47, 142]. We used storyboards for three main reasons: i. they allow to elicit responses from a diverse sample of respondents as they provide a common visual language; ii. they enable respondents to focus on the essential interaction design and visual elements of the mechanism and reduce the noise caused by other unrelated design elements; iii. when combined with surveys, they enable researchers to collect large volumes of data, thus enabling statistical analysis. Note that while our survey focused on both photos and videos, our storyboards focused only on photos as they are easier to visualize. While the research method used in this study allowed us to collect answers from a representative sample of Internet users, we note also that the collected data provides speculative opinions formed on the basis of storyboards. It does not tell us how people would behave if the presented methodologies were implemented in the OSNs they use, nor how well or poorly these designs would perform in practice.

4.2 Design of the Survey Instrument

The questionnaire contained 71 items. The full questionnaire is reported in Supplementary Material A (pages 1-8).[8] The questions were organized into four sections. The first section contained demographic questions used to ensure that the sample was representative of the US adult Internet population (i.e., Q1 to Q4). It also contained standardized questions to assess the respondents' socio-economic status (SES) [43], and two questions focused on OSN to collect information about the platforms and technologies most used by the respondents. The second section focused on MPCs. The questions were designed to ask respondents to alternatively take the role of an *uploader* (i.e., the person causing the MPC) and that of a *data subject* (i.e., the person suffering from the MPC). To design the questions, we loosely followed the critical incident method [33]. The questions focused on the most recent events where respondents experienced an MPC.[9] We designed the questions to be semi-structured to avoid priming respondents and to enable respondents to freely report the ways in which they prevent/resolve MPCs. The third section focused on severe episodes resulting from MPCs. We focused, in particular, on episodes of discrimination and exposure suggested by recent literature [25]. Specifically, the section contained questions on revenge pornography [134, 137, 146], public shaming [75, 149], and discrimination [46]. In the final section, we presented respondents with storyboards for each of the mechanisms discussed in Section 3 (storyboards are depicted in Supplementary Material B, pages 9-13).

The storyboards were drawn following the design guidelines by Truong et al. [142]. The sequences in the storyboard were drawn following a short narrative drafted iteratively and reviewed by the authors of this paper. The short stories depicted the same character sharing a photo showing multiple individuals on a social network. Each sequence also showed how the system would dissuade or prevent the uploader from sharing without the consent of the involved individuals. After each storyboard was presented, the respondents were shown three seven-point Likert scales for rating the desirability of the mechanism as an uploader and as a data subject. We ensured that respondents understood the mechanism depicted in each storyboard by asking them two comprehension questions: i. to describe in their own words what was happening in each storyboard, and ii. to list the pros and cons of the depicted mechanism.

4.3 Procedure

The survey was approved by the IRB of the University of Lausanne. We contracted a vendor to deploy and conduct the survey; the vendor was in charge of selecting a representative sample of adult Internet users based in the US [32].[10] As the vendor did not provide a system for collecting the consent of legal guardians or parents for younger individuals, we recruited only adult respondents. The panelists were recruited via partnerships and invited via banners and messaging, and then went through quality controls.

To eliminate possible presentation and carryover effects, questions presenting multiple choices, as well as the order of the storyboards, were randomized. Before deploying the questionnaire, we conducted eight pre-tests involving individuals at our institution. One of the authors sat with the participants and, for each question, asked the participant to re-state, in their own words, what the question was asking and how they would answer. Feedback provided at this stage was used to adjust wording (e.g., added the qualifier *concrete* for Q17) and provide additional context (e.g.,

[8] See https://doi.org/10.17605/OSF.IO/NDBK7, last accessed January 2021.

[9] As the data is self-reported, it depends on whether the respondent noticed the MPC and how they perceived it.

[10] We chose to rely on web panels, through a vendor, in order for our results to generalize to adult Internet users in the US including 50+ y.o. users who usually cannot be reached through crowd-sourcing platforms (e.g., MTurk). See the work of Redmiles et al. [119] for a thorough comparison of the generalizability of security and privacy survey results for different subject pools (incl. crowd-sourcing and web panels).

added an informative text that precedes Q27). A second pre-test was conducted via *soft launch* with a random sample of 100 respondents following the same deployment procedure for the *full launch*. The answers provided in the open questions helped us assess whether respondents understood the questions. In some cases, we added qualifiers and informative text to clarify the intent (e.g., added *privacy reasons* to Q15). The soft launch data was not used in the analysis. Last, it took an average of 17 minutes to complete the questionnaire. Upon completion of the questionnaire, each respondent received ~USD 5 (average hourly pay: ~USD 18).[11]

4.4 Factor Analyses using Regression Models

To gain better insight into factors that influence respondents' preference towards specific mechanisms, we estimated the parameters in mixed-effect regression models: M1 for "as an uploader" and M2 for "as a data subject". Independent variables include demographic variables, the frequency of sharing on social networks, whether they caused or suffered an MPC, the level of concern about exposing/being exposed, and whether they caused or suffered a severe MPC. A random intercept accounts for the respondent effect. The fixed effects were selected using an AIC-based procedure.

In addition, as we were interested in whether respondents assigned different desirability values to precautionary and dissuasive mechanisms, we computed two additional regression models: M3 (as an uploader) and M4 (as a data subject). These models included interaction terms between the observed variables and the mechanism types. For readability, we reduced the mechanism types to two levels: precautionary (CS, FB, and N) and dissuasive (SS, E, AL, PS, and LT). Lastly, to better understand whether specific types of users have a preference towards any of the dissuasive mechanisms, we built another model M5 that includes the interaction effects between the observed variables and five mechanisms (SS, E, AL, PS, and LT). For readability, Model M5 was computed only for "as a data subject" as the results were similar to "as uploader". Furthermore, M5 was restricted to the dissuasive mechanisms, as the precautionary mechanisms had higher rankings and did not differ from each other.

All the models were fitted using the `lme4` and `lmerTest` packages of the statistical software `R` (see [14]). The tests are corrected for multiple comparisons using the one-step method implemented in the `multcomp` package.[12] Although, with such a large amount of data, the tests validity is guaranteed by the asymptotic theory, we were mindful not to over-interpret results and p-values (see for example Pinheiro and Bates [109]). Note that we refrained from using Friedman tests to compare the mechanisms since we wanted to control for the participant factors like gender, age, etc. Due to the number of tests and the complexity of the statistical analyses, it was not possible to accurately compute a sample size. However, we performed an approximate power analysis of the ANOVA F-test (on 8 groups with $\alpha = .05$). We estimated 1800 respondents to reach a power of 90% to detect a "small" scale effect equal to 0.1. Therefore, we requested from our vendor 1900 complete responses after the low-quality responses had already been removed.

4.5 Data Reliability and Coding Process

To ensure data reliability, several quality-assurance (QA) processes were followed when administering the survey instrument: *speeders* and *straightliners* were removed at the source by the vendor. After removing speeders and straightliners, 1907 responses were collected. Upon inspection of the answers to the open-ended questions, we further excluded 115 respondents, which left us with $N = 1792$ valid responses. To analyze Q42 (i.e., an open-ended question about pros and cons of

[11] We cannot provide the exact amount as the vendor uses an incentive scale which is based on panelist characteristics.
[12] To support reproducibility, the `R` scripts are available at https://doi.org/10.17605/OSF.IO/NDBK7, last accessed January 2021.

each mechanism), we used thematic analysis [19]. Thematic analysis has been used by earlier work in privacy research [63] to analyze open-ended survey questions. One of the authors manually pre-processed the dataset and removed respondents who clearly did not provide reliable responses (e.g., "*great thingsrutghrniow*", "*none none none*"). As a result, 6, 552 comments were selected. The filtered data were iteratively read (for data familiarization) and coded, where both semantic and latent features of data were considered. We built a codebook inductively, while keeping in mind relevant privacy theories, including concepts such as privacy boundaries (particularly MPCs as potential boundary turbulence) [107] and acceptable information flows depending on the context [96]. A total of 3, 246 anecdotes were coded, and a code list including 102 codes was generated. Later the codes were clustered into potential themes given their relevance on how respondents expressed their perceived benefits and concerns about different mechanisms. The initial themes were discussed by authors using a thematic map and these themes were either merged or removed to develop the final list of 17 main themes (i.e., 6 for benefits and 11 for concerns, reported in Sec. 5.3 and 5.4).

4.6 Demographics and General Statistics

The proportion of female respondents was 51.8% (or 928 respondents) and the age distribution was as follows: 18-29 (22.7%, or 406 respondents), 30-39 (15.3%, or 274), 40-49 (19.1%, or 342), 50-59 (18.1%, or 325), 60+ (24.8%, or 445). Respondents were well-distributed across the four macro regions of the US: 22.1% (or 396) live in the Midwest, 20.1% (or 360) live in the Northeast, 33.8% (or 606) live in the South, and 24.0% (or 430) live in the West. Regarding sharing photos and videos,[13] the majority of the respondents (62.9%, or 1127) use OSN (e.g., Facebook, Tumblr, Twitter, etc.); about 35.7% (or 639) use IM apps (e.g., Messenger, WeChat, Snapchat, etc.); and 13.9% (or 249) use multimedia sharing websites (e.g., Flickr, Youtube, Vimeo, etc.).

4.6.1 Multimedia Sharing Practices, MPCs, and Coping Strategies. Most respondents declared sharing online content where others are shown (84.82% or 1520) and/or appearing in online content shared by others (85.66% or 1535), showing the potential of MPCs. In fact, many respondents experienced an MPC in the last 12 months, with 16.29% (or 292) causing it, and 24.39% (or 437) suffering it.[14] The reasons for MPC were similar to previous work (e.g., [138]): individuals shown looked drunk, the content showed some nudity, or they did not like their appearance (cf. Table 1 of Supplementary Material C, page 14 for details).

Importantly, a significant association was found between "Suffered MPC" and "Gender" ($\chi^2(1) = 9.64$, $p < .01$), with "Female-Suffered" being over-represented. Similarly, a significant association was found between "Age" and "Suffered MPC", and "Age" and "Caused MPC" (resp. $\chi^2(4) = 65.44$, $p < .001$, $\chi^2(4) = 81.2$, $p < .001$), with "18-29-Caused/Suffered" being over-represented. This suggests that younger users most often reported dealing with MPC (both causing/suffering), and that female respondents were more likely to report having suffered an MPC (cf. Table 2 (a) and (c) of Supplementary Material C, page 14 for details).

Regarding the coping strategies users employ to try to avoid or resolve MPCs when they appear, the results confirm previous research [77, 116, 117, 138] (cf. Table 3 of Supplementary Material C, page 15 for details). Strategies to *avoid* MPCs included: asking permission first from individuals shown in the content (37%), not sharing content (22%), using common sense (18%), sharing

[13] Respondents could select multiple categories.

[14] These numbers could be a lower-bound of the MPC phenomena, because of the qualifying text 'for privacy reasons' in the question, which might have led multiple respondents to only report incidents they perceive privacy-related. In previous studies without this qualifier, such as [138], participants reported a higher proportion of what, after analysis, was actually deemed to be an MPC.

Table 1. Ratings of the precautionary and dissuasive mechanisms reported by the respondents of the survey. The ranks are computed from the mean desirability.

	As uploader				*As data subject*			
	Rank	**Mean**	**SD**	**95% C.I.**	**Rank**	**Mean**	**SD**	**95% C.I.**
CS (prec.)	*1*	4.68	1.68	[4.61; 4.76]	*2*	4.75	1.70	[4.68; 4.83]
FB (prec.)	*2*	4.68	1.73	[4.60; 4.76]	*1*	4.76	1.73	[4.68; 4.84]
N (prec.)	*3*	4.67	1.67	[4.59; 4.75]	*3*	4.73	1.67	[4.65; 4.81]
LT (diss.)	*4*	4.54	1.85	[4.45; 4.62]	*4*	4.63	1.86	[4.54; 4.71]
PS (diss.)	*5*	4.43	1.84	[4.35; 4.52]	*5*	4.50	1.87	[4.42; 4.59]
AL (diss.)	*7*	4.39	1.91	[4.30; 4.48]	*6*	4.46	1.91	[4.38; 4.55]
SS (diss.)	*8*	4.11	1.93	[4.02; 4.20]	*8*	4.16	1.94	[4.07; 4.25]
E (diss.)	*6*	4.39	1.76	[4.31; 4.47]	*7*	4.45	1.77	[4.37; 4.53]

only personal media (9%), or sharing with a restricted group only (3%). *Resolution* strategies included: deleting the content (53%), apologizing (17%), discussing with the offended party (6.5%), self-reflection to avoid repeating the mistake (9.7%), or no action (10.5%). In most cases these solutions are ineffective to avoid conflict. They might even make conflicting content known to others or worse, copied and redistributed. This further reinforces the need to address MPC *before* the content is shared online [77, 116, 138], and justifies the need for precautionary and/or dissuasive mechanisms.

4.6.2 MPC consequences and severity. Regarding MPC consequences, respondents stated (cf. Table 4 of Supplementary Material C, page 15 for a full listing): a breach of personal privacy (22.45%), discrimination (20.79%), hurting others' feelings (16.31%), their reputation (7.59%), or no clear consequence (17.02%). Regarding MPC severity, we found that 12.05% of respondents caused severe MPCs to other individuals. The breakdown by type revealed that 64.90% were conflicts related to attire, facial expressions or alcohol use with public shaming, and 35.10% were due to nudity or sexual content. We also found that 7.37% of respondents experienced severe MPCs caused by others, among which 25.42% for content related to nudity or sex. This finding suggests that about 1 out of 10 adult Internet users in the US experienced severe MPC that might have led to public shaming, discrimination or revenge pornography. In terms of significant associations, we only found that younger users caused and suffered severe MPC more than older users (resp. $\chi^2(4) = 95.85$, $p < .001$, $\chi^2(4) = 63.29$, $p < .001$). See Table 2 (b) and (d) of Supplementary Material C for details.

5 MAIN FINDINGS

For conciseness and clarity, we will focus on and report the details of statistically significant results, and discuss only non-significant results that are relevant. The complete tables, including non-significant findings can be found in Supplementary Materials.

5.1 Mechanism Desirability as Uploader and Data Subject

Respondents compared the different storyboards that represent the mechanisms. These were rated as uploaders and as data subjects. The Likert tableau describing the ratings of the mechanisms is reported in Figure 4. Table 1 shows the average ratings of the mechanisms and their ranks. Significant differences between mechanism desirability ratings as "an uploader" were found (Friedman test: $\chi^2(7) = 218.71$, $p < .001$). No difference was found between the precautionary mechanisms (i.e., CS, FB, and N). However, the precautionary mechanisms ratings appeared higher than those of the dissuasive mechanisms (except the pair N vs. LT). Among the dissuasive mechanisms, AL,

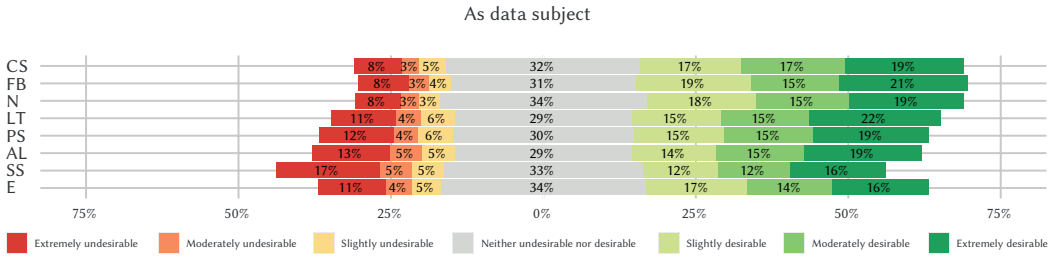


Fig. 4. Likert tableau reporting the ratings of the different mechanisms compared in the survey (as data subject).

Table 2. Regression model coefficients for the desirability value respondents associated with the different strategies presented in the survey. M1 describes the desirability as *uploader*, while M2 the desirability as *data subject*. For readability, Just few coefficients are shown below. The full table can be found in Supplementary Material E, page 17, Table 1.

Model	**M1: des. as uploader**				**M2: des. as data subject**			
Factors	*Est.*	*Std Err*	*z-value*	*p-value*	*Est.*	*Std Err*	*z-value*	*p-value*
(intercept) Des. Value	3.92	.20	19.65	<.001	3.98	.20	19.76	<.001
Gender : Male	-.16	.07	-2.32	<.05	-.19	.07	-2.69	<.01
Age : 60+	.47	.12	3.79	<.001	.51	.12	4.11	<.001
Freq. being shown : High	.27	.07	3.62	<.001	.25	.07	3.41	<.001
Level of Care : High	.36	.13	2.76	<.01	.34	.13	2.61	<.01
Level of Concern : High	.44	.08	5.55	<.001	.48	.08	5.98	<.001
Caused Severe MPC : Yes	.31	.10	2.98	<.01	.34	.10	3.26	<.01

PS and E did not differ, and LT was not different from PS. All other comparisons lead to statistically significant differences. The post-hoc tests adjusted p-values are reported in Table 5 of Supplementary Material D, page 16. Rating the mechanisms as "a data subject" revealed significant differences between the mechanisms ($\chi^2(7) = 249.76$, $p < .001$). The post-hoc analysis showed similar results to "as an uploader" (see Table 5 of Supplementary Material D).

We also did a simple analysis of how respondents changed their desirability scores when they moved from "data subject" to "uploader" (Supplementary Material E, page 17). The plots revealed that around 80% of respondents did not change their scores. Moreover, 20% of respondents who changed their decisions, did so in both directions (i.e., increasing and decreasing the scores). Thus, there is no clear trend between respondents' role and desirability of the mechanisms.

5.2 Factors that Influence the Desirability of the Mechanisms

Table 2 shows some coefficients for the desirability value associated with the mechanisms as uploader (M1) and as data subject (M2). Full results are in Supplementary Material F, pages 18-33. Figure 5 presents a visual summary. We now review the factors that remained significant after model selection. It is worth mentioning that the size of the effects, measured here by the estimates, corresponds to the difference in the desirability value between the reference level and the considered level (e.g., *Female* versus *Male* for *Gender*). It is an arbitrary task to define when such a difference is large, medium or small, beyond their significance. In the analysis below, we qualified any effect smaller than 0.2 in absolute value, as *small*, from 0.2 to 0.5 as *medium*, and larger than

0.5 as *large*.[15] When the effect is small, we added some words or caution on over-interpreting the results.

Gender. The gender of the respondents influences the desirability level (Tables 7 and 8 of Supplementary Material F). **Women rated *any* mechanisms significantly higher** than men[16] (M1: Est. −.16, small effect, Z=−2.32, $p < .05$, M2: Est. −.19, small effect, Z=−2.69, $p < .05$). Comparing precautionary and dissuasive, no significant different rating was found between male and female, though, with a mild trend, females assigned higher values to precautionary over dissuasive mechanisms. Following prior research stating they are the target of gender-based violence with higher frequency [30], it is reasonable that women value *any* solution to control attacks higher. Further, women might prefer precautionary mechanisms, as these offer more control over dissuasive ones. It also appeared that women prefer the mechanism with 'individual' consequences rather than 'social' ones (M5: "PS - LT": Est. −.20, Z=−2.03, $p < .05$, Table 33 of Supplementary Material F). While women rated all dissuasive mechanisms higher than men, they did not do the same for Public Shaming (PS). Free-texts from female respondents revealed that warnings with social consequences were perceived to **further expose the privacy of the data subjects**: "*This is making private information public. If Alice reported a picture and it was removed, why are you telling everyone about it. This makes the problem even worse*" [F, 50-59 y.o.]. This is in line with prior reports [31, 125] observing that women usually do not speak about harassment in public to protect their social reputation, and avoid blame. However, these findings should be taken with caution and may be subject to further investigations considering the generally low estimates of the effects.

Age. We found a significant effect of age on the desirability level of different mechanisms (Tables 9 and 10 of Supplementary Material F). As uploaders and as data subjects, **older respondents (60+) rated any mechanism higher** than younger ones (M1: 18-29: Est. .47, medium effect, Z=3.79, $p < .001$, 30-39: Est. .35, medium effect, Z=2.78, $p < .05$; M2: 18-29: Est. .51, large effect, Z=4.11, $p < .001$, 30-39: Est. .39, medium effect, Z=3.02, $p < .05$). As uploaders, precautionary and dissuasive mechanisms are rated differently by respondents in different age groups (M3: $F = 10.5$, $df_1 = 4$, $df_2 = 12530$, $p < .001$). Globally, precautionary mechanisms are preferred, though the difference is larger for younger ages (e.g., "Diss. − Prec." estimated at −.374 for 18-29) and smaller for older ages (e.g., "Diss. − Prec." estimated at −.025 for 60+). For the oldest ages, difference of precautionary and dissuasive is not significant. Similar results were obtained when respondents rated the mechanisms as data subjects (M4: $F = 14.3$, $df_1 = 4$, $df_2 = 12530$, $p < .001$). Lastly, in line with previous models, among dissuasive mechanisms, older respondents (60+) rated all mechanisms higher than younger ones (e.g., M5: 18-29,E: Est. −.30, $Z = -2.28$, $p < .05$, Table 34 of Supplementary Material F).

This suggests a gap between generations: younger adults, at ease with social networks, and older adults, struggling with new IT tools [12]. While younger users prefer control, **older users prefer the simpler interaction flows** afforded by the dissuasive mechanisms: "*Seems to me this [ConsenShare] is a lot of work to do just to show a picture. People should not be taking inappropriate pictures*" [F, 60+ y.o.]; "*It [ConsenShare] requires a lot of time and hassle*" [F, 60+ y.o.]. Lastly, it is worth noting that like any statistical analysis, it cannot exclude confounding factors with age that could cause variation of the rating. One could think of a "frequency of sharing online" that is only partially controlled by the factor "frequency of being shown".

[15] We defined these levels given that most of the desirability scores were cast in the upper part of the scale, namely 3 to 7 (global mean 4.52, global sd 1.81). Therefore, an increase of 0.2 in the estimate corresponds to a 5% increase in desirability, and an increase of 0.5 in the estimate corresponds to a 12.5% increase in desirability.

[16] Though non-binary gender was allowed (Q1), no gender other than female or male was recorded.

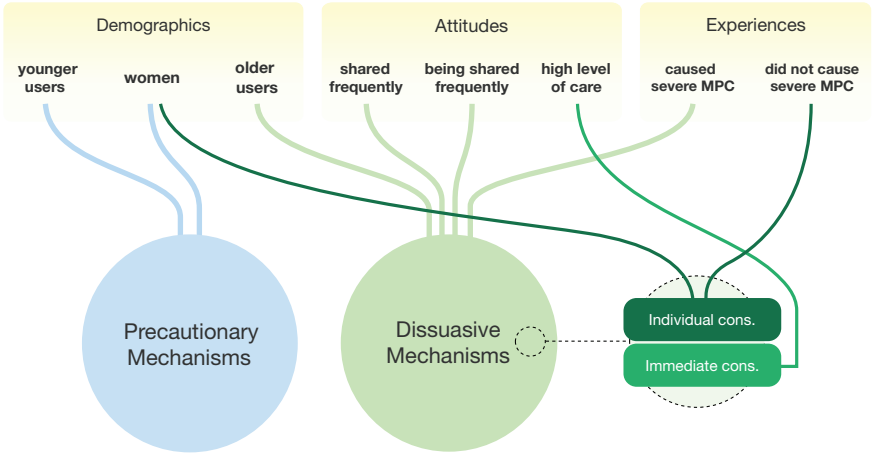


Fig. 5. Summary of the factor analysis. Lines in the figure represent significant associations between the demographic profile, the attitudes, the experiences of the user and the mechanisms tested in this study. The methods on the far right represent a subset of the dissuasive mechanisms.

Frequency of sharing or being shared online. For conciseness we use "share" to refer to uploaders and "being shared" to refer to data subjects. **Respondents who shared intensively rated *any* mechanism higher** than those who shared rarely or never (M1: Est. .27, medium effect, Z=3.62, $p < .01$). Similar results are true for data subjects, those appearing in content being shared (M2: Est. .25, medium effect, Z=3.41, $p < .01$). See Tables 13 and 14 in Supplementary Material F. As uploaders, there was no significant difference of the ratings of precautionary and dissuasive mechanisms between categories. However, we note a trend: the estimate of the difference between precautionary and dissuasive mechanisms narrows down to zero as the frequency of sharing online increases (M3: Est.= $-.28$, $-.25$, $-.16$, resp. for levels=low, mid, high). Similar results are true for the frequency of being shared. See Tables 23 and 29 in Supplementary Material F.

This suggests that, as the frequency of sharing/being shared increases, the preference towards precautionary mechanisms decreases. A possible explanation suggested by the free-text comments is that **users who reported to share frequently consider the additional effort** that precautionary mechanisms might add to the sharing process: "*It is a whole hassle to upload a picture [using FaceBlock] and wait for consent*" [shares daily, F, 18-29 y.o.]; "*It [ConsenShare] may take a lot more time to post a single picture*" [shares daily, M, 40-49 y.o.]. Similarly, those whose contents get shared intensively, might dislike having to approve every content where they appear: "*It [ConsenShare] would result in extra notifications for me to check on; I would find it [Negotiation] a hassle to have to be constantly monitoring these things.*" [being shared daily, F, 40-49 y.o.].

Care about exposing others and concern about being exposed. Users with self-reported **high level of care about exposing others rated *any* mechanism higher** than users who have a self-reported mid level of care (M1: Est. .45, medium effect, Z=4.69, $p < .001$) and than users with self-reported low level of care (M1: Est. .36, medium effect, Z=2.77, $p < .05$). Similarly, users with self-reported high level of care about being exposed rated any mechanism higher than users with self-reported mid level of care (M2: Est. .33, medium effect, Z=3.88, $p < .001$) and than users with self-reported low level of care (M2: Est. .48, medium effect, Z=5.98, $p < .001$). See Tables 15 to 18 in Supplementary Material F.

For the 'care about exposing', there is no significant difference in the ratings between categories. No clear trend appeared. For the 'concern of being exposed', there are significant differences in the ratings between categories (M3: $F = 13.8$, $df_1 = 2$, $df_2 = 12530$, $p < .001$). The preferred mechanisms are the precautionary ones especially for Concern=*Low* (M3: Est.= $-.412$, $Z = -7.12$, $p < .001$). For the two highest levels of concern the difference between precautionary and dissuasive mechanisms is smaller (M3: Mid: Est= $-.110$, $Z = -1.86$, $p = .06$, and High: Est= $-.162$, $Z = -3.11$, $p < .01$). Results from model M4 are similar in terms of the significance of the interaction terms (M4: Care: $F = 3.51$, $df_1 = 2$, $df_2 = 12530$, $p = .03$; Concern: $F = 12.0$, $df_1 = 2$, $df_2 = 12530$, $p < .001$). See Tables 24-25 and 30-31 in Supplementary Material F.

In short, these findings reveal that **users weigh-in their attitudes towards MPCs when evaluating the mechanisms**. Users with self-reported high level of care tend to value *any* mechanism providing protection from MPCs. Results are inconclusive whether these users value precautionary or dissuasive mechanisms more.

Further investigation showed that respondents with self-reported high level of care about exposing others rated solutions with "immediate" consequences (i.e., LT, PS) significantly higher compared respondents with self-reported mid and low level of care (M5: "LT, Low - High": Est. $-.39$, $Z=-2.19$, $p < .05$, "LT, Mid - High": Est. $-.51$, $Z=-3.85$, $p < .01$, "PS, Low - High": Est. $-.37$, $Z=-2.07$, $p < .05$, "PS, Mid - High": Est. $-.53$, $Z=-3.98$, $p < .01$, Table 38 of Supplementary Material F). The analysis of the free-text comments revealed that respondents with high level of care felt LT and PS provided responses that were **proportionate to the crime** (i.e., sharing without consent): "*I like this [Law Threat] because it is more factual and more appropriately relays the seriousness of the offense*" [high level of care, F, 60+ y.o.]; "*This [Public Shaming] would cause the user [uploader] to feel the same, or worse, amount of shame that the target [data subject] felt.*" [high level of care, F, 18-29 y.o.]; "*This [AL] is way too harsh of a punishment*" [slightly careful, student, F, 18-29 y.o.]; "*The penalty [SS] is just way out of the ballpark and it does not match the severity of the offense*" [high level of care, F, 30-39 y.o.].

Causing or Suffering from a severe MPC. As uploaders, the users who caused a severe MPC rated *any* mechanism higher than users who did not (M1: Est. .31, medium effect, $Z=2.98$, $p < .01$, see Table 19 in Supplementary Material F). Furthermore, there was no significant difference in the rating assigned to precautionary and dissuasive mechanisms between respondents who did and who did not cause a severe MPC. Thus, **the personal history of the respondents plays a role in the ratings of each mechanism**. To avoid being again in stressful conditions, users prefer any safety mechanism preventing trouble.

Among dissuasive mechanisms, respondents who did not cause a severe MPC disliked mechanisms with "social" consequences (i.e., AL and PS) compared to Law Threat (M5: "Did not cause, AL - LT": Est. $-.15$, $Z=-2.02$, $p < .05$, "Did not cause, PS - LT": Est. $-.13$, $Z=-1.69$, $p = .09$, Table 40 of Supplementary Material F). The free-text analysis showed how these users try to **avoid repeating the mistake of the uploader**: "*It just opens up a can of worms by shaming someone who might have innocently posted a photo. If that isn't the case and it was done in a vindictive manner then the legal aspect would be more appropriate.*" [did not cause severe MPC, F, 60+ y.o.]. Interestingly, suffering a severe MPC was not found significant in any model. That may be due to collinearity between caused and suffered severe MPC. One may guess that people causing and suffering severe MPCs are in fact the same. However, this hypothesis needs further investigation.

5.3 Perceived Benefits Reported by Respondents

The thematic analysis performed on the comments left by respondents on each mechanisms revealed six main reasons that explain the overall positive ratings.[17] About a quarter of the analyzed comments described all of the mechanisms as a **necessity for social media**. Respondents described how *any* mechanism that could better protect the privacy of the OSN users is beneficial. Particularly, respondents recognized how precautionary mechanisms provide the user of OSN more control over the images being shared, or the audience who has access to them: "*Social media users would be able to dictate exactly how they want to be depicted in any photo that is uploaded to that site, which would create a much safer and less exploited online reality.*" [M, 30-39 y.o.]. Respondents also appreciated how all mechanisms intervene before sharing thus potentially reducing the overall number of conflicts: "*The pro is I am pre warned that I may possibly humiliate or hurt a friend or family. I also get a chance to think about what I am doing before it is too late.*" [F, 18-29 y.o.]. All of the presented methods could also potentially alert the data subjects of non-consensual uploads and of pictures that concerned them that could be circulating on social media: "*The pro is being able to discover when you are being put online by others that you may not be aware of and having the choice to give consent or not.*" [F, 30-39 y.o.]. Many respondents also highlighted how these methods are particularly beneficial for vulnerable subjects, particularly *younger users*, who might approach social media naively and expose themselves in the process: "*This is a good way to prevent things like this from happening especially for young teens who may be totally unaware and naive about how these things can cause torment to others lives for years if posted.*" [F, 60+ y.o.].

5.3.1 Perceived Benefits of Precautionary mechanisms. Several respondents favoured the precautionary mechanisms as these where considered **more democratic** than the dissuasive mechanisms. Respondents appreciated that with precautionary mechanisms every person represented in the picture had a say on whether the photo was appropriate for being shared online, not just the uploader: "*I think this is a great solution as it allows the other person to have a say in what happens with the photo and allows for a wider variety of outcomes.*" [F, 18-29 y.o.]. Respondents appreciated precautionary strategies as able to provide control to the data subjects over the content shared online (and to data subjects caught unintentionally in the shared content): "*It gives more control to the people involved whether or not they want to be a public shared photo/video. Also it helps filter out the people in the background that were caught in the photo/video unintentionally*" [M, 30-39 y.o.]. The Negotiation method was appreciated because it enforced the collaboration between the uploader and the data subject, thus soliciting peers to reach consensus: "*Both sides can decide on what they like on the sharing preferences for the photo. This allows the 'conversation' to not be one-sided.*" [F, 18-29 y.o.]. One of the said advantage of precautionary mechanisms over other methods was that these did not restrict the ability of OSN users to share content but they simultaneously enabled data subjects to **vet the appropriateness of the content** being shared: "*I like that it can leave in the people who consented to the picture, while blurring out the people who didn't. This allows the person the freedom to post their picture, just without the person who didn't want to be shown.*" [M, 30-39 y.o.].

5.3.2 Perceived Benefits of Dissuasive mechanisms. Many comments concerned dissuasive mechanisms. Respondents appreciated these methods because they can encourage uploaders to **think twice** to avoid negligence: "*Sometimes people post stuff without thinking. A gentle reminder of what they are doing and the possible consequences is all that's needed for most people*" [F, 50-59 y.o.]; "*This would be more up my alley to prevent a person from doing something silly in the heat of the moment.*" [M, 50-59 y.o.]. Dissuasive mechanisms were also seen as **strong enough to deter** people from

[17] Themes are marked in **bold**.

sharing without consent: "*I like how strict this warning is, and that it gives the person who wants to send the picture out 'sufficient warning' that not only will their account be penalized, but that their monetary standing will be jeopardized.*" [F, 60+ y.o.]. Several respondents saw **dissuasive mechanisms more appropriate for severe MPCs**. Several comments described how these mechanisms can be helpful when the shared content contains nudity: "*I like this idea if someone is actually trying to put nude pictures of someone or porn without consent the person portrayed in the photo should be able to take action on the person that did it.*" [F, 50-59 y.o.]; and for cyberbullying: "*It prevent someone from posting mean and hateful things which could prevent things like suicide.*" [M, 40-49 y.o.]. Also, given that several dissuasive methods make the identity of the offender known to the peers, they were considered effective for serial offenders: "*The pros is that this technique could stop serial abusers and prevent them from posting yet again, without the other people's consent.*" [M, 40-49 y.o.]. We now look at the concerns raised by the respondents.

5.4 Concerns Reported by Respondents

In addition to the benefits that most respondents reported in the survey, many comments discussed concerns that respondents identified in the storyboards. A large portion of these concerns applied equally to precautionary and dissuasive methods, and we organized them in five themes.

Most respondents pointed out that a critical aspect of the reviewed methods is related to **consent collection**. These methods need to record and store trace of the explicit consent given by the data subject(s) appearing in the shared media. Several respondents pointed out how providing consent offline would not be protective for the uploader, as data subjects can later change their mind and deny having approved the publication of the content: "*The major con of this is people being in good relationships turn bad and the person given consent could easily changed their mind and lie.*" [M, 18-29 y.o.]. Furthermore, respondents pointed out that these methods would generally put a burden on the uploaders because it might be difficult to get consent for every shared media that shows other people, for pictures of large groups of people, or getting consent for media that portray peoples that are hard or impossible to reach (e.g., children, people without a OSN account, people who might have died after the photo was taken, celebrities): "*Some people like my grandma are never going on Facebook so I'll never get her consent and all my pictures of Grandma are going to be ruined.*" [F, 18-29 y.o.]. Finally, respondents also indicated how these methods would also put a burden on the data subjects because they have to explicitly provide consent to all media being shared where they are displayed. Particularly, providing consent would be annoying for people that are often portrayed in social media posts (e.g., social proxies, celebrities): "*How would this work with celebrities or famous people? Would they be able to approve every photo posted of them. With regular people I think it will work fine but I'm curious how it would work for public figures.*" [F, 18-29 y.o.].

Respondents also described how any method for fighting MPC would need to be protective of **vulnerable data subjects and uploaders**. In fact, data subjects who report a non-consensual sharing might be retaliated by uploaders if their identity would become known: "*This can lead to harassment of exposed user from the user that uploaded the photo.*" [F, 18-29 y.o.]; "*I think they should use fictitious names for safety or assign a number to users.*" [F, 60+ y.o.]. Several respondents also described how these mechanisms that are designed to protect data subjects might be gamed to maliciously ostracize and discriminate uploaders: "*Too easy for malicious, vindictive, or angry people to make others lives harder in a few moments. To easy to whip up the outrage mob and say 'see, this person got banned! They must be a bad Person, you need no other evidence!'.*" [F, 30-39 y.o.].

Another theme identified in the analysis is that these methods **may limit the freedom of speech**. Participants described how improper or excessive appeal to deterrence or explicit consent collection might be abused to limit journalistic work: "*It's easy to imagine this being used to censor*

legitimate, newsworthy content. Imagine if a public figure tried to erase an unwanted image or video from the Internet even if that content is worthwhile and newsworthy." [M, 40-49 y.o.]. Felons might use these mechanisms to hide information that might otherwise be necessary to the justice system: "*What if someone records a crime or a police officer doing something wrong but this technique blocks them from posting it? That could cause serious problems and might even prevent justice from being done by preventing videos recorded without consent from being used as evidence in criminal cases.*" [M, 30-39 y.o.].

Furthermore, respondents highlighted how these methods can be **circumvented by switching to a different sharing platform** that does not implement the same policies: "*This would not solve the problem since someone could just use another site to post photos if they were inclined to do.*" [F, 60+ y.o.]. Finally, several respondents had concerns regarding the **reliability of the technology** powering these methods. Specifically, respondents were concerned that the machine learning algorithms required for recognizing people in the shared media might under- or over-trigger: "*The tech behind it sounds amazing, but in actuality facial recognition systems seem a bit unforgiving. I imagine the algorithm would need altering too, say if plastic surgery is involved or drastic losing weight.*" [F, 18-29 y.o.].

5.4.1 Concerns with Precautionary mechanisms. In addition to the general concerns described above, respondents also described concerns specific to precautionary mechanisms, that we organized in three themes. Many comments described how mechanisms such as **FaceBlock and ConsenShare would disrupt the user experience** of OSN users. Respondents described how the sharing flows would require additional effort and become confusing for many users: "*Too many steps to go through and it ruins the entire experience of sharing a photo.*" [F, 50-59 y.o.]. Furthermore, comments described how pixelating or masking parts of the content would significantly degrade the aesthetics of images or videos shared on social media: "*The pixelation seems a little far fetched and a bit awkward, especially when someone doesn't get around to approving of the photo being posted of themself.*" [F, 18-29 y.o.].

A second theme specific to precautionary methods concerned the **privacy risks**. Many respondents did not trust 3rd party companies to handle their privacy on OSNs as these could track their behaviors across several networks: "*The con is you have an additional website gathering personal data about you, your friends or associates, about your online behavior and decision making process.*" [F, 30-39 y.o.]. Most of the respondents were also uncomfortable sharing their ID card with OSNs for proving their identity: "*This is giving social media sites not only a selfie but a picture of your identification card to use as they want. The stated goal will be to keep your identity from being shared, the unstated goal would be to mass-collect information on an ID card for millions of people and it will all be kept, analyzed, stored, and shared with business partners.*" [M, 18-29 y.o.]. Respondents also pointed out that OSNs are not immune to security problems. Their servers could be hacked and the original un-blurred content stored on the server might be leaked by hackers: "*Whatever site you're uploading to that has this feature, still has access to your original photo.*" [M, 18-29 y.o.]; "*What if the social network platform got hacked. People's information will get leaked easily.*" [M, 30-39 y.o.].

Finally, the way precautionary methods **hid identities** was seen insufficient. Several respondents explained how face obfuscation might not be always effective as the person despite the blurring might still be recognized by contextual factors (e.g., clothes, tattoos, background, friends, comments left by other peers): "*If people know you well enough and know your inner circle of friends, you're still going to get caught or pointed out that the blurred person is you ... people enjoy to do witch hunts or guessing games at who is the person that's been hidden...*" [F, 18-29 y.o.].

5.4.2 Concerns with Dissuasive mechanisms. A large body of comments left by respondents concerned dissuasive methods. Their analysis helps also understand the ratings given by respondents.

We organized these in three themes. Many respondents highlighted that the consequences of dissuasive methods are **disproportionate** to the committed crime: "*This would provide too steep of a consequence and would allow important things like your ability to get a loan affected by things that should not affect it. I don't like this at all. The positive is that because the consequence is so steep people will be really careful as to what they share*" [M, 30-39 y.o.]. Expectedly, many respondents felt the Social Score mechanism would clash with the libertarian culture of the U.S. and believed these methods should assign penalties that should be limited to the social network only: "*I'm not about to be involved in some social credit score commie crap! I'm an American. Do this by the laws of America, not China!*" [M, 40-49 y.o.]; "*Decreasing your chances of getting a loan has nothing to do with your social media presence and should not be affected by what you post.*" [F, 18-29 y.o.].

Several respondents also explained their ratings saying that dissuasive mechanisms would only provide **partial prevention** from MPCs. Many felt these methods would be ineffective against uploaders who intentionally might want to harm others, or adolescents who might not appreciate the consequences of triggering the penalties, or other types of users who might not care about financial or social consequences of their misbehavior: "*I feel that the likely outcome would be that person who is going to post an objectionable photo is beyond being swayed by an empathic warning and would still go ahead and post the photo out of spite or hatred ... this warning would just be ignored by most objectionable photo posters and it wouldn't deter them at all.*" [F, 60+ y.o.]. Additionally, respondents recognized that these methods might become less effective over time due to habituation: "*I think that after the first few times the novelty will wear off and it wouldn't be as effective and be more of an annoyance.*" [F, 18-29 y.o.].

Finally, several respondents did not appreciate Public Shaming because this method responded to a wrongdoing (i.e., the MPC) with **retaliation**, which many considered also wrong: "*Public shaming is as bad as posting the picture, and it is too late after the picture has already been posted. Two wrongs don't make a right.*" [F, 60+ y.o.]. Many respondent also found that this method could also easily backfire by attracting more attention to the MPC thus further exposing the data subjects: "*If Alice reported a picture and it was removed, why are you telling everyone about it. This makes the problem even worse. Now people who didn't see the picture are going to be asking about it.*" [F, 50-59 y.o.]. In the next section, we turn to discuss our main findings. We also revisit our research hypothesis and discuss the study implications.

6 DISCUSSION

The results reported in this article, as discussed next, contribute to the first and third research streams described in the literature review section (cf. Sec. 2).

6.1 MPCs are Serious and Widespread Threats

Concerning the first research stream (cf. Sec. 2.1), the results presented in Section 4.6, beyond confirming previous work, shed more light into the incidence and severity of MPCs and on particular groups of users. According to our results, 24.39% of respondents suffer MPCs as data subjects, while 16.29% cause MPCs as uploaders. Unfortunately, for 7.37% of the respondents some of these conflicts lead to serious consequences such as public shaming, discrimination or revenge porn. The data collected with the survey also helped to identify women and younger users as the most vulnerable user profiles. In particular, women reported suffering MPCs with higher frequency than men (cf. Section 4.6.1). Similarly, younger users cause and suffer MPC (and severe MPC) with higher frequency than older users. These results might explain why both younger users and women preferred precautionary mechanisms, where they might get a better protection against privacy violations. Lastly, these findings suggest that social media platforms should take steps to proactively assist data subjects who suffer from MPC to report abuse, and address uploaders' behavior.

6.2 Precautionary Mechanisms Are Preferred

The core finding of this research contributes to the third research stream discussed in Section 2.3: our respondents prefer the precautionary mechanisms over the dissuasive ones.

6.2.1 Precautionary Mechanisms Are Preferred by Both Uploaders and Data Subjects. The prevalence of the ratings given to precautionary mechanisms did not change even when we asked respondents to rate the desirability of the techniques putting themselves in the shoes of uploaders and data subjects. While the preference as data subject matched our expectation that respondents would prefer more control, we did not expect the same result for uploaders. When the person was in charge of the process (i.e., being an uploader), we expected a preference towards more manageable solutions with simpler workflows. Surprisingly, we did not observe this difference. An explanation suggested by the thematic analysis is that respondents prefer mechanisms that remove the uncertainty that uploaders might often face when deciding whether content involving multiple parties is appropriate for being shared (cf. 5.3.1). The results of the regression model indicate that women and younger users assign higher desirability ratings to precautionary mechanisms. In general, the analysis of the survey responses revealed that precautionary methods were perceived to provide more control and protection to OSN users. These results **confirm H1**, as precautionary methods *enforce* collaboration, allow more control, and lower uncertainty around sharing media portraying multiple persons on OSNs.

6.2.2 No Differences Between Precautionary Mechanisms. In terms of the types of precautionary mechanisms and whether respondents preferred any of them, we did not find any statistically significant difference. That is, respondents considered precautionary mechanisms to be more desirable than dissuasive mechanisms, but they did not seem to prefer precautionary mechanisms that modify the content (FB and CS) over precautionary mechanisms that modify the audience (N). Further research should be conducted in terms of ascertaining whether precautionary mechanisms that modify the content and those that modify the audience could be perceived as more or less beneficial or desirable in practice. One could argue that both approaches are complementary and mechanisms that combine both approaches would be worth considering and testing for their acceptability in the future. For instance, one could only share a photo without modification with a subset of the audience and with modifications (e.g., face blurred) with the rest of the audience, or keep faces blurred until an appropriate audience has been agreed upon. Lastly, in terms of item modification mechanisms (FB and CS), no preference emerged between those that would imply giving their ID to the social media provider (FB) over those that would rely on a trusted third party for this (CS). The thematic analysis revealed that several respondents did not trust 3rd party companies to handle their privacy (cf. Sec. 5.4.1). Likely, these were counterbalanced by many other respondents who, on the other hand, were uncomfortable sharing their ID card with OSNs.[18]

6.3 Challenges and Opportunities of Dissuasive Mechanisms

Dissuasive mechanisms might reduce conflicts caused by mindless sharing. Considering the factor analyses, we identified different profiles of users who might benefit from dissuasive mechanisms. In addition, we found dissuasive mechanisms consistently received lower ratings than precautionary ones. We highlight four facets of the factor analysis that help explain these results.

6.3.1 Law Threat Is Preferred. Our findings showed, as we hypothesized, that mechanisms that threaten users with immediate consequences (LT and PS) received significantly higher approval compared with those that contain messages with delayed consequences (SS and AL). This is in line

[18]Note that ID card verification is used by several popular services (e.g., `Airbnb`). In practice, once the identity of the user has been verified, there is no need to keep the scan of the identification document.

with previous research that found that immediate punishments achieve higher compliance [1, 51]. Particularly, respondents perceived LT as the most desirable dissuasive mechanism. Respondents noted that referring to the law was the most sensible choice in most situations. In contrast, the comments received about SS illustrate that many did not perceive the punishment commensurate to the crime and clashing with the libertarian culture of the US. Therefore, these results only **partially confirm H2**: while respondents reported higher desirability for the LT among the dissuasive methods, we did not measure differences in desirability among the precautionary methods (cf. Sec. 6.2.2). Furthermore, the results allow us to **confirm H3**, as respondents indicated higher desirability for methods with immediate punishments.

6.3.2 Effectiveness of Empathy. The design of the Empathy was appreciated by some as it provided uploaders a chance to "think twice" about how others appearing in the multimedia content could feel if that content were to be released to others. However, most respondents rated this as one of the least effective mechanisms. We studied this mechanism because prior work on smoking cessation [127, 128] and education [100] found that empathy-arousing messages could persuade users to follow healthier behaviors. However, more recent work in the privacy domain reveals that OSN users might not decrease the likelihood of sharing content when primed to consider the *privacy of others* [8]. Consistently, the desirability ratings collected in our study show that priming users to consider the *consequences* of non-consensual sharing was not considered effective by respondents. We also found that E does not significantly differ from AL and PS (shame arousing mechanisms). This would suggest that empathy anticipation yields a similar effect to the shame anticipation. In a sense, one might feel ashamed for not empathizing with others. Further research is required to understand the relation between empathy and shame anticipations in this context.

6.3.3 Frequent Sharers and Older Users Prefer Dissuasive Mechanisms. The results of the regression model suggest that potential users consider the trade-offs between a feature's usefulness, their level of technical expertise, and the frequency with which they share multimedia content online when assigning a desirability value to mechanisms that counter MPCs. Users that frequently share multimedia content online seek faster publication flows. Older users seek the simplest interaction mechanism. For these two categories of users, dissuasive mechanisms afforded more desirable solutions. Going forward mechanisms to counter MPCs should be tested specifically with frequent sharers and older users to ensure these minimize the cost over the publication workflow and that these have a fast learnability. The results of the regression model also suggest that women and younger users expressed higher desirability for precautionary mechanisms (cf. Sec. 6.2.1). Taken together, these results only **partially confirm H4**: while we did not find difference in desirability based on the role (i.e., uploader or data-subjects), we found difference in the desirability ratings for specific groups of respondents based on gender, age, frequency of sharing, and MPC experience.

6.3.4 Dissuasive Mechanisms and Habituation. Some of the respondents did not trust dissuasive mechanisms could be effective in the long run. Several comments pointed out that over time users could simply become inattentive to the warnings (i.e., habituation effect), hence these could lose their dissuasive properties. Habituation effects on security warnings have been studied extensively in prior work (for a review see the work of [9]). Earlier approaches used different techniques to regain user attention to the warnings. Two main approaches were applied: i. *dishabituation* [11, 20, 21] where every time a stimulus takes a 'new' form the user is not familiar with; ii. *spontaneous recovery* [144] where user attention is recovered if the time gap between two warnings is large enough (i.e., no stimulus exposure for some time). The first mechanism has been extensively studied as it offers many design opportunities. For instance, we could imagine techniques

that might involve iterative modification of size and color (text, background), use of flashy borders, changing the location of the warnings, and the way that warnings pop up, similar to what is typically done for ad banner blindness [113]. Last, given equivalent desirability ratings of some of the dissuasive mechanisms, we might alternate the use of mechanisms with similar effectiveness.

6.4 Implications for Future Design

Our analysis helped identify the following implications that aim to reduce the occurrence of MPCs.

6.4.1 Preventing Side Leakages from Modified Content. Mechanisms relying on content modification (such as FB and CS) would not guarantee that an MPC might not occur. Through inline or separate commenting systems peers that are close to the individuals who are shown in the content might leak information on who is behind the cutouts, thus generating privacy conflict. The identity of data subjects can also be inferred based on their clothes or the location where the photo was taken. Several respondents expressed this concern (cf. Sec. 5.4.1). Two possible solutions are suggested. The first would be to block comments on the media until all the involved parties have given their consent and the parts that are masked are completely removed. A second, safer, approach would be to withhold the publication of the material until all involved parties had given consent. We acknowledge that these solutions will complicate or delay the publication flow. To reduce the impact we can imagine to perform automated content analysis (cf. 6.4.2) and trigger these mechanisms only for sensitive content.

6.4.2 Consent Collection and Appeal Process. Several respondents highlighted that dissuasive mechanisms were lacking the possibility to demand reconsideration in case data subjects flagged uploaded content as non-consensual. The design we tested did not enable the uploaders to present evidence of prior conversations where approval had been sought and/or obtained. This additional feature would be necessary to ensure that dissuasive mechanisms would not be exploited by data subject to purposely cause damage to the uploaders. This also calls for accountable consent collection (cf. the work of Olteanu et al. [101]). What if the data subject says yes and reports the MPC later? How can the uploader collect legal proof that the data subject agreed to share the content? [19] Furthermore, several respondents noted that the tested mechanisms will add additional burden on the uploaders to collect consent (cf. Sec. 5.4). Given the amount of photo sharing, social media users might never resort to obtaining explicit consent for every single photo. Many photos, due to their sensitivity or who are on them, are rather unlikely to cause problems. Going forward, we might combine these mechanisms with existing work on automated content analysis [102, 122] and automated privacy risk detection [129, 156], which could be used to determine whether to trigger privacy protection mechanisms similar to those tested in this research.

6.4.3 Censorship, Journalistic Work, and Freedom of Speech. Many respondents pointed out that whatever the mechanism put in place to control MPCs, this should not limit journalistic work or freedom of speech in general. Precautionary mechanisms, in particular, could be exploited by public figures to control the publication of material by citizens and journalists that could be of public interest. This discussion is similar to the debate around "the right to be forgotten" that was raised a few years ago for search engines [126]. From an algorithmic point of view, it would be extremely complicated to automatically allow or deny publication of content as the notion of public interest would need to be better specified. A likely simpler approach to the matter might be to flag specific accounts as belonging to public figures and apply different policies to these. Alternatively, we could imagine crowdsourcing this effort, possibly concerning the point discussed above

[19] As an extreme example of this principle, we refer the reader to the app `Legal Fling`. See https://legalfling.io, last retrieved March 2020.

(cf. Sec. 6.4.2). Note also that social media platforms are nowadays called to exert more control over what is published online (e.g., to reduce fake news, fight extremism). However, these platforms need to moderate without incurring into censorship [54]. In most cases these platforms use *algorithms* to decide the appropriateness of content. These programs might often over trigger, thus limiting the publication of legitimate uploads (see the example of Facebook blocking posts displaying art pieces because of nudity [40, 123]). We argue that ultimately decision-making algorithms should take into account users' opinions on the legitimacy and appropriateness of content.

7 LIMITATIONS, FUTURE WORK, AND CONCLUSIONS

Limitations. This study relies on self-reported rather than measured behaviors. It is conceivable that respondents may misreport their own preferences and behavior either to self-justify or because of recall bias. This could also be due to the well-known privacy paradox [97] where respondents may behave differently from their attitudes. This is a limitation to any survey study (especially in privacy research [3]) not specific to our study. We found negligible differences between the desirability ratings respondents provided as uploaders and as data-subjects. However, this result might have been biased by the use of a single storyboard to represent both roles. Furthermore, testing multiple scenarios through a large-scale survey would have been complex. This could have been useful to study whether preference towards mechanisms would change when the same picture would have been taken in different contexts (e.g., party vs. work dinner). We recognize context is highly subjective and hard to render in a storyboard (not to mention to model computationally). In the study, we use the presence of multiple faces as a *trigger* to flag media that could potentially lead to privacy conflicts and then we let depicted people *decide* on whether there is a privacy violation based on both the content and on the context.

Future Work. As a next step, we are planning to implement some of these mechanisms in an OSN and to conduct a longitudinal AB test. This design will enable us to compare objective behavioral data on the number of reported MPC incidents in a real-usage scenario. On the other hand, we explored dissuasive mechanisms aiming to arouse three types of emotions including fear, shame, and empathy. Moving forward, other types of emotions might be studied to fight MPCs. Furthermore, future studies involving objective data should study ways to counter habituation effects. Finally, in the future it would be relevant to consider nudges in the context of MPCs (cf. Sec. 2.4 and Sec. 3.1.2).

Conclusions. The paper contributes to the research area of multiparty privacy by analyzing the risks from MPCs, and by testing novel mechanisms and comparing them to existing ones to help users manage MPCs. The large-scale survey revealed that a substantial portion of US adult Internet users experience MPCs and that this leads to severe consequences (e.g., public shaming). Given the lack of proper privacy-preserving tools in current OSNs, the compared mechanisms were highly relevant and well received by the respondents of the survey. This study provides a user-centric evaluation of the desirability of these methods and a discussion of the advantages and disadvantages of the different approaches. In general, respondents (including women and young adults) preferred precautionary mechanisms, as they provide more control to the data subjects and allow uploaders to reduce the uncertainty about what is appropriate for sharing. However, frequent sharers and older users seemed to favour dissuasive mechanisms, as they offered faster and simpler interactions. The study revealed a number of limitations of the tested designs that will need to be addressed in derivative design and follow-up research. To conclude, our study provides valuable insights for privacy designers, social scientists, and policy makers, aiming at raising interest and awareness in the community for MPCs. Given that these events might have serious consequences

in the life of social media users, more research is needed to study how to best protect users from privacy conflicts generated by sharing multimedia content online.

ACKNOWLEDGMENTS

This work was partially funded by the Swiss National Science Foundation with Grant #CRSK-2_190762. We sincerely thank William Delamare, Lahari Goswami, Mathias Humbert, Francesca Mosca, Alexandra-Mihaela Olteanu and Mael Péquignot for providing precious feedback on early versions of this article. We also thank Holly Cogliati, James Tyler, and Vincent Vandersluis for proofreading this article. Last, we thank all the respondents of the survey.

REFERENCES

[1] Ann J. Abramowitz and Susan G. O'Leary. 1990. Effectiveness of delayed punishment in an applied setting. *Behavior Therapy* 21, 2 (March 1990), 231–239. https://doi.org/10.1016/S0005-7894(05)80279-5

[2] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (Jan. 2015), 509–514. https://doi.org/10.1126/science.aaa1465

[3] Alessandro Acquisti and Jens Grossklags. 2004. Privacy Attitudes and Privacy Behavior. In *Economics of Info. Secu.* Vol. 12. Kluwer Academic Publishers, Boston, MA, USA, 165–178. https://doi.org/10.1007/1-4020-8090-5_13

[4] Alessandro Acquisti, Manya Sleeper, Yang Wang, Shomir Wilson, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, and Florian Schaub. 2017. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Comp. Surv. (CSUR)* 50, 3 (Aug 2017), 1–41. https://doi.org/10.1145/3054926

[5] Paarijaat Aditya, Rijurekha Sen, Peter Druschel, Seong Joon Oh, Rodrigo Benenson, Mario Fritz, Bernt Schiele, Bobby Bhattacharjee, and Tong Tong Wu. 2016. I-Pic: A Platform for Privacy-Compliant Image Capture. In *Proc. of the Annual Int. Conf. on Mobile Sys., Applications, and Services (MobiSys)*. ACM, Singapore, Singapore, 235–248. https://doi.org/10.1145/2906388.2906412

[6] Gulsum Akkuzu, Benjamin Aziz, and Mo Adda. 2020. Towards consensus-based group decision making for co-owned data sharing in online social networks. *IEEE Access* 8 (2020), 91311–91325. https://doi.org/10.1109/ACCESS.2020.2994408

[7] Irwin. Altman. 1975. *The environment and social behavior: privacy, personal space, territory, crowding*. Brooks/Cole Pub. Co., Monterey, California, USA. https://eric.ed.gov/?id=ED131515

[8] Mary Jean Amon, Rakibul Hasan, Kurt Hugenberg, Bennett I Bertenthal, and Apu Kapadia. 2020. Influencing Photo Sharing Decisions on Social Media: A Case of Paradoxical Findings. In *2020 IEEE Symp. on Secu. and Privacy (SP)*. IEEE, San Francisco, California, USA, 79–95. https://doi.org/10.1109/SP40000.2020.00006

[9] Ammar Amran, Zarul Fitri Zaaba, and Manmeet Kaur Mahinderjit Singh. 2018. Habituation effects in computer security warning. *Info. Secu. Jour.: A Global Perspective* 27, 2 (Mar 2018), 119–131. https://doi.org/10.1080/19393555.2018.1448492

[10] Reza Anaraky, Tahereh Nabizadeh, Bart Knijnenburg, and Marten Risius. 2018. Reducing Default and Framing Effects in Privacy Decision-Making. In *SIGHCI 2018 Proc.* Assoc. for Info. Sys. (AIS), Atlanta, GA, USA, 7. https://aisel.aisnet.org/sighci2018/19

[11] Bonnie Brinton Anderson, C. Brock Kirwan, Jeffrey L. Jenkins, David Eargle, Seth Howard, and Anthony Vance. 2015. How Polymorphic Warnings Reduce Habituation in the Brain: Insights from an FMRI Study. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'15)*. ACM, New York, NY, USA, 2883–2892. https://doi.org/10.1145/2702123.2702322

[12] Anne Aula. 2005. User study on older adults' use of the Web and search engines. *Universal Access in the Info. Society* 4, 1 (Sep 2005), 67–81. https://doi.org/10.1007/s10209-004-0097-7

[13] Solon Barocas and Karen Levy. 2019. Privacy Dependencies. *Wash. Law Rev.* 95, ID 3447384 (Sept. 2019), 62. https://papers.ssrn.com/abstract=3447384

[14] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Jour. of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[15] Roy F. Baumeister and Jill Lobbestael. 2011. Emotions and antisocial behavior. *The Jour. of Forensic Psychiatry & Psychology* 22, 5 (Oct. 2011), 635–649. https://doi.org/10.1080/14789949.2011.617535 Publisher: Routledge.

[16] Filipe Beato and Roel Peeters. 2014. Collaborative joint content sharing for online social networks. In *Proc. of the Int. Conf. on Pervasive Comp. and Comm. Workshops (PERCOM Workshops)*. IEEE, Budapest, Hungary, 616–621. https://doi.org/10.1109/PerComW.2014.6815277

[17] Andrew Besmer and Heather Richter Lipford. 2010. Moving beyond untagging: photo privacy in a tagged world. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'10)*. ACM, Atlanta, GA, USA, 1563. https://doi.org/10.1145/1753326.1753560

[18] Rachel Botsman. 2017. Big data meets Big Brother as China moves to rate its citizens. https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion

[19] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[20] Cristian Bravo-Lillo, Saranga Komanduri, Lorrie Faith Cranor, Robert W. Reeder, Manya Sleeper, Julie Downs, and Stuart Schechter. 2013. Your Attention Please: Designing Security-Decision UIs to Make Genuine Risks Harder to Ignore. In *Proc. of the Symp. on Usable Priv. and Secu. (SOUPS'13)*. ACM, New York, NY, USA, Article Article 6, 12 pages. https://doi.org/10.1145/2501604.2501610

[21] José Carlos Brustoloni and Ricardo Villamarín-Salomón. 2007. Improving Security Decisions with Polymorphic and Audited Dialogs. In *Proc. of the Symp. on Usable Priv. and Secu. (SOUPS'07)*. ACM, New York, NY, USA, 76–85. https://doi.org/10.1145/1280680.1280691

[22] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'19)*. ACM, New York, NY, USA, Article Paper 503, 15 pages. https://doi.org/10.1145/3290605.3300733

[23] Barbara Carminati and Elena Ferrari. 2011. Collaborative Access Control in On-line Social Networks. In *Proc. of the Int. Conf. on Collaborative Comp.: Networking, Applications and Worksharing (CollaborateCom'11)*. IEEE, Orlando, FL, USA, 231–240. https://doi.org/10.4108/icst.collaboratecom.2011.247109

[24] Hichang Cho and Anna Filippova. 2016. Networked Privacy Management in Facebook: A Mixed-Methods and Multinational Study. In *Proc. of the ACM Conf. on Comp.-Supported Cooperative Work & Social Comp. (CSCW'16)*. ACM, San Francisco, CA, USA, 502–513. https://doi.org/10.1145/2818048.2819996

[25] Ben C. F. Choi, Zhenhui (Jack) Jiang, Bo Xiao, and Sung S. Kim. 2015. Embarrassing Exposures in Online Social Networks: An Integrated Perspective of Privacy Invasion and Relationship Bonding. *Info. Sys. Research* 26, 4 (Dec 2015), 675–694. https://doi.org/10.1287/isre.2015.0602

[26] Tharntip Tawnie Chutikulrungsee and Oliver Kisalay Burmeiste. 2017. Interdependent Privacy. *ORBIT Jour.* 1, 2 (Oct 2017), 14. https://doi.org/10.29297/orbit.v1i2.38

[27] Anna L. Cox, Sandy J.J. Gould, Marta E. Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proc. of the ACM Conf. Extended Abstracts on Human Factors in Comp. Sys. (CHI EA'16)*. ACM, New York, NY, USA, 1389–1397. https://doi.org/10.1145/2851581.2892410

[28] Simeon de Brouwer. 2020. Privacy Self-Management and the Issue of Privacy Externalities: of Thwarted Expectations, and Harmful Exploitation. *Internet Policy Rev.* 9, 4 (Dec. 2020), 29. https://doi.org/10.14763/2020.4.1537

[29] Istituto Poligrafico e Zecca dello Stato S.p.A. 2019. Gazzetta Ufficiale. https://www.gazzettaufficiale.it/eli/id/2019/07/25/19G00076/sg

[30] European Union and Agency for Fundamental Rights. 2015. *Violence against women: an EU-wide survey : main results.* Publications Office, Luxembourg, Luxembourg. https://doi.org/10.2811/981927

[31] Chai Feldblum and Victoria Lipnic. 2016. Select Task Force on the Study of Harassment in the Workplace (Full Report). https://www.eeoc.gov/eeoc/task_force/harassment/upload/report.pdf

[32] Thom File and Camille Ryan. 2013. *Computer and Internet Use in the United States.* Number ACS-28 in American Community Surv. Reports. US Census Bureau Washington, DC, Washington, DC, USA. https://www.census.gov/library/publications/2014/acs/acs-28.html

[33] John C. Flanagan. 1954. The Critical Incident Technique. *Psychological Bulletin* 51, 4 (Jul 1954), 327–358. https://www.apa.org/PsycINFO/cit-article.pdf

[34] Arturo Flores and Serge Belongie. 2010. Removing pedestrians from Google street view images. In *2010 IEEE Comp. Society Conf. on Comp. Vision and Pattern Recognition - Workshops*. IEEE, New York, NY, USA, 53–58. https://doi.org/10.1109/CVPRW.2010.5543255 ISSN: 2160-7516.

[35] Ricard Fogues, Jose M. Such, Agustin Espinosa, and Ana Garcia-Fornes. 2015. Open Challenges in Relationship-Based Privacy Mechanisms for Social Network Services. *Int. Jour. of Human-Comp. Interaction (CHI)* 31, 5 (May 2015), 350–370. https://doi.org/10.1080/10447318.2014.1001300

[36] Ricard L. Fogues, Pradeep K. Murukannaiah, Jose M. Such, and Munindar P. Singh. 2017. Sharing Policies in Multiuser Privacy Scenarios: Incorporating Context, Preferences, and Arguments in Decision Making. *ACM Trans. on Comp.-Human Interaction (TOCHI)* 24, 1 (Mar 2017), 1–29. https://doi.org/10.1145/3038920

[37] Ricard L. Fogues, Pradeep K. Murukannaiah, Jose M. Such, and Munindar P. Singh. 2017. SoSharP: Recommending Sharing Policies in Multiuser Privacy Scenarios. *IEEE Internet Comp.* 21, 6 (Nov 2017), 28–36. https://doi.org/10.1109/MIC.2017.4180836

[38] Ricard L Fogues, Pradeep Murukanniah, Jose M Such, Agustin Espinosa, Ana Garcia-Fornes, and Munindar Singh. 2015. Argumentation for multi-party privacy management. In *Proc. of the Int. Workshop on Agents and CyberSecu. (ACySe'15)*. ACM Press, Istanbul, Turkey, 3–6. https://eprints.lancs.ac.uk/id/eprint/74191/

[39] Secretariat for Legal Affairs. 1969. American Convention on Human Rights. http://www.oas.org/dil/treaties_B-32_American_Convention_on_Human_Rights.htm Last accessed 2nd of October 2020.

[40] Agence France-Presse. 2018. Facebook apologises for censoring prehistoric Venus statue. https://phys.org/news/2018-03-facebook-apologises-censoring-prehistoric-venus.html Last accessed 6th of October 2020.

[41] Augusta Gaspar and Mariana Henriques. 2018. Driven by shame : How a negative emotion may lead to prosocial behaviour. In *New Interdisciplinary Landscapes in Morality and Emotion*, Sara Graça Da Silva (Ed.). Taylor and Francis Group, London, UK, 53–66. https://doi.org/10.4324/9781315143897-5

[42] Jack P Gibbs. 1968. Crime, punishment, and deterrence. *The Southwestern Social Science Quarterly* 48, 4 (1968), 515–530. http://www.jstor.org/stable/42867909

[43] Dennis L. Gilbert. 2008. *The American class structure in an age of growing inequality* (7th ed ed.). Pine Forge Press, Los Angeles, CA, USA. https://www.goodreads.com/book/show/4107424-the-american-class-structure-in-an-age-of-growing-inequality

[44] Gouvernement.fr. 2019. Protect Your Image Rights. https://www.gouvernement.fr/guide-victimes/en-proteger-son-droit-a-l-image Last accessed 2nd of October 2020.

[45] Patricia Greenspan. 1995. *Practical Guilt: Moral Dilemmas, Emotions, and Social Norms*. Oxford University Press, London, United Kingdom. https://doi.org/10.2307/2653776

[46] Ralph Gross and Alessandro Acquisti. 2005. Information revelation and privacy in online social networks. In *Proc. of the ACM Workshop on Priv. in the Electronic Society (WPES'05)*. ACM, Alexandria, VA, USA, 71–80. https://doi.org/10.1145/1102199.1102214

[47] Mieke Haesen, Jan Meskens, Kris Luyten, and Karin Coninx. 2010. Draw Me a Storyboard: Incorporating Principles & Techniques of Comics.... In *Proc. of the 24th BCS Interaction Specialist Group Conf. (BCS'10)*. BCS Learning & Development Ltd., Swindon, GBR, 133–142. https://dl.acm.org/doi/10.5555/2146303.2146323

[48] Rakibul Hasan, David Crandall, Mario Fritz, and Apu Kapadia. 2020. Automatically Detecting Bystanders in Photos to Reduce Privacy Risks. In *IEEE Symp. on Secu. and Priv. (S&P'20)*. IEEE, Oakland, CA, USA, 318–335. https://doi.org/10.1109/SP40000.2020.00097

[49] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer Experience of Obscuring Scene Elements in Photos to Enhance Privacy. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'18)*. Assoc. for Comp. Mach. (ACM), New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173621

[50] Rakibul Hasan, Yifang Li, Eman Hassan, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. 2019. Can Privacy Be Satisfying? On Improving Viewer Satisfaction for Privacy-Enhanced Photos Using Aesthetic Transforms. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'19)*. Assoc. for Comp. Mach. (ACM), New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300597

[51] Yusuke Hayashi, Anne M. Foreman, Jonathan E. Friedel, and Oliver Wirth. 2019. Threat appeals reduce impulsive decision making associated with texting while driving: A behavioral economic approach. *PLOS ONE* 14, 3 (March 2019), e0213453. https://doi.org/10.1371/journal.pone.0213453 Publisher: Public Library of Science.

[52] Steven D Hazelwood and Sarah Koon-Magnin. 2013. Cyber Stalking and Cyber Harassment Legislation in the United States: A Qualitative Analysis. *Int. Jour. of Cyber Criminology* 7, 2 (2013), 14. https://www.cybercrimejournal.com/hazelwoodkoonmagninijcc2013vol7issue2.pdf

[53] Benjamin Henne, Christian Szongott, and Matthew Smith. 2013. SnapMe if you can: privacy threats of other peoples' geo-tagged media and what we can do about it. In *Proc. of the ACM Conf. on Secu. and Priv. in Wireless and Mobile Networks (WiSec'13)*. ACM, Budapest, Hungary, 95. https://doi.org/10.1145/2462096.2462113

[54] Jesse Hirsh. 2019. Why Social Platforms Are Taking Some Responsibility for Content. https://www.cigionline.org/articles/why-social-platforms-are-taking-some-responsibility-content Last accessed 12th of October 2020.

[55] Roberto Hoyle, Luke Stark, Qatrunnada Ismail, David Crandall, Apu Kapadia, and Denise Anthony. 2020. Privacy Norms and Preferences for Photos Posted Online. *ACM Trans. Comput.-Hum. Interact.* 0, ja (2020), 1. https://doi.org/10.1145/3380960

[56] Hongxin Hu and Gail-Joon Ahn. 2011. Multiparty Authorization Framework for Data Sharing in Online Social Networks. In *Proc. of the IFIP Annual Conf. on Data and Applications Secu. and Priv. (DBSec'11)*. Springer, Richmond, VA, USA, 29–43. https://doi.org/10.1007/978-3-642-22348-8_5

[57] Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. 2011. Detecting and resolving privacy conflicts for collaborative data sharing in online social networks. In *Proc. of the Annual Comp. Secu. Applications Conf. (ACSAC'11)*. ACM, Orlando, FL, USA, 103–112. https://doi.org/10.1145/2076732.2076747

[58] Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. 2012. Enabling Collaborative data sharing in Google+. In *Proc. of the IEEE Global Comm. Conf. (GLOBECOM'12)*. IEEE, Anaheim, CA, USA, 720–725. https://doi.org/10.1109/GLOCOM.2012.6503198

[59] Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. 2013. Multiparty Access Control for Online Social Networks: Model and Mechanisms. *IEEE Trans. on Knowledge and Data Engineering* 25, 7 (Jul 2013), 1614–1627. https://doi.org/10.1109/TKDE.2012.97

[60] Mathias Humbert, Benjamin Trubert, and Kévin Huguenin. 2019. A Survey on Interdependent Privacy. *ACM Comp. Surv.* 52, 6, Article Article 122 (Oct. 2019), 40 pages. https://doi.org/10.1145/3360498

[61] Panagiotis Ilia, Barbara Carminati, Elena Ferrari, Paraskevi Fragopoulou, and Sotiris Ioannidis. 2017. SAMPAC: Socially-Aware Collaborative Multi-Party Access Control. In *Proc. of the ACM on Conf. on Data and Application Secu. and Priv. (CODASPY'17)*. ACM, Scottsdale, AZ, USA, 71–82. https://doi.org/10.1145/3029806.3029834

[62] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. 2015. Face/Off: Preventing Privacy Leakage From Photos in Social Networks. In *Proc. of the ACM SIGSAC Conf. on Comp. and Comm. Secu. (CCS'15)*. ACM, Denver, CO, USA, 781–792. https://doi.org/10.1145/2810103.2813603

[63] Timo Jakobi, Sameer Patil, Dave Randall, Gunnar Stevens, and Volker Wulf. 2019. It Is About What They Could Do with the Data: A User Perspective on Privacy in Smart Metering. *ACM Trans. Comput.-Hum. Interact.* 26, 1, Article 2 (Jan. 2019), 44 pages. https://doi.org/10.1145/3281444

[64] Jeffrey Rubin and Dana Chisnell. 2011. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests, 2nd Edition | Wiley* (2nd edition ed.). Wiley, Hoboken, NJ, USA. https://books.google.ch/books/about/Handbook_of_Usability_Testing.html?id=l_e1MmVzMb0C&redir_esc=y

[65] Eric J. Johnson and Daniel Goldstein. 2003. Do Defaults Save Lives? *Science* 302, 5649 (2003), 1338–1339. https://doi.org/10.1126/science.1091721 arXiv:https://science.sciencemag.org/content/302/5649/1338.full.pdf

[66] Adam N. Joinson. 2008. Looking at, Looking Up or Keeping Up with People?: Motives and Use of Facebook. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'08)*. ACM, Florence, Italy, 1027. https://doi.org/10.1145/1357054.1357213

[67] Evans Jonathan and Frankish Keith. 2009. *In Two Minds: Dual Processes and Beyond - Open Research Online*. Oxford University Press, London, UK. 382 pages. http://oro.open.ac.uk/22096/

[68] Daniel Kahneman. 2011. Thinking, fast and slow (Kindle Edition). https://doi.org/10.1037/h0099210

[69] Dilara Kekulluoglu, Nadin Kokciyan, and Pinar Yolum. 2018. Preserving Privacy as Social Responsibility in Online Social Networks. *ACM Trans. on Internet Tech. (TOIT)* 18, 4 (Apr 2018), 1–22. https://doi.org/10.1145/3158373

[70] Dilara Keküllüoğlu, Nadin Kökciyan, and Pınar Yolum. 2016. Strategies for Privacy Negotiation in Online Social Networks. In *Proc. of the Int. Workshop on AI for Priv. and Secu. (PrAISe)*. ACM, The Hague, The Netherlands, 1–8. https://doi.org/10.1145/2970030.2970035

[71] Simon Kemp. 2020. Digital 2020: Global Digital Overview. https://datareportal.com/reports/digital-2020-global-digital-overview Last accessed 13th of November 2020.

[72] Nadin Kökciyan, Nefise Yaglikci, and Pinar Yolum. 2017. An Argumentation Approach for Resolving Privacy Disputes in Online Social Networks. *ACM Trans. Internet Tech.* 17, 3, Article 27 (June 2017), 22 pages. https://doi.org/10.1145/3003434

[73] Genia Kostka. 2019. China's social credit systems and public opinion: Explaining high levels of approval. *New Media & Society* 21, 7 (2019), 1565–1593. https://doi.org/10.1177/1461444819826402

[74] Bert P. Krages. 2003. The Photographer's Right. http://www.krages.com/phoright.htm Last accessed 2nd of October 2020.

[75] Justin Kruger, Cameron L. Gordon, and Jeff Kuban. 2006. Intentions in teasing: When "just kidding" just isn't good enough. *Jour. of Personality and Social Psychology* 90, 3 (Mar 2006), 412–425. https://doi.org/10.1037/0022-3514.90.3.412

[76] Airi Lampinen. 2015. Networked Privacy beyond the Individual: Four Perspectives to "Sharing". In *Proc. of The Fifth Decennial Aarhus Conf. on Critical Alternatives (CA'15)*. Aarhus University Press, Aarhus N, 25–28. https://doi.org/10.7146/aahcc.v1i1.21300

[77] Airi Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. 2011. We're in it together: interpersonal management of disclosure in social network services. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'11)*. ACM, Vancouver, BC, Canada, 3217–3226. https://doi.org/10.1145/1978942.1979420

[78] Yifang Li, Nishant Vishwamitra, Hongxin Hu, and Kelly Caine. 2020. Towards A Taxonomy of Content Sensitivity and Sharing Preferences for Photos. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'20)*. Assoc. for Comp. Mach. (ACM), New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376498

[79] Yifang Li, Nishant Vishwamitra, Bart P. Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Effectiveness and Users' Experience of Obfuscation as a Privacy-Enhancing Technology for Sharing Photos. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 67 (Dec. 2017), 24 pages. https://doi.org/10.1145/3134702

[80] Kaitai Liang, Joseph K. Liu, Rongxing Lu, and Duncan S. Wong. 2015. Privacy Concerns for Photo Sharing in Online Social Networks. *IEEE Internet Comp.* 19, 2 (Mar 2015), 58–63. https://doi.org/10.1109/MIC.2014.107

[81] Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt. 2019. Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'19)*. ACM, New York, NY, USA, 1–18. https://doi.org/10.1145/3290605.3300361

[82] Karolina Mania. 2020. The Legal Implications and Remedies Concerning Revenge Porn and Fake Porn: A Common Law Perspective. *Sexuality & Culture* (May 2020), 19. https://doi.org/10.1007/s12119-020-09738-0

[83] Hiroaki Masaki, Kengo Shibata, Shui Hoshino, Takahiro Ishihama, Nagayuki Saito, and Koji Yatani. 2020. Exploring Nudge Designs to Help Adolescent SNS Users Avoid Privacy and Safety Threats. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'20)*. Assoc. for Comp. Mach. (ACM), New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376666

[84] Shigenori Matsui. 2015. The Criminalization of Revenge Porn in Japan. *Wash. Int. Law Jour.* 24, 2 (2015), 289–318. https://heinonline.org/HOL/P?h=hein.journals/pacrimlp24&i=305

[85] Claude-Hélène Mayer and Elisabeth Vanderheiden. 2019. *Bright Side Of Shame: transforming and growing through practical.* Springer, Cham, Switzerland. http://link.springer.com/10.1007/978-3-030-13409-9

[86] Barbara A. Mellers and A. Peter McGraw. 2001. Anticipated Emotions as Guides to Choice. *Current Directions in Psychological Science* 10, 6 (Dec. 2001), 210–214. https://doi.org/10.1111/1467-8721.00151 Publisher: SAGE Publications Inc.

[87] Terance D Miethe, Hong Lu, et al. 2005. *Punishment: A comparative historical perspective.* Cambridge University Press, Cambridge, UK. https://books.google.ch/books/about/Punishment.html?id=o2ovr4ZzIXsC&redir_esc=y

[88] Andrew D. Miller and W. Keith Edwards. 2007. Give and take: a study of consumer photo-sharing culture and practice. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'07)*. ACM Press, San Jose, California, USA, 347–356. https://doi.org/10.1145/1240624.1240682

[89] Tehila Minkus, Kelvin Liu, and Keith W Ross. 2015. Children seen but not heard: When parents compromise children's online privacy. In *Proc. of the 24th Int. Conf. on World Wide Web (WWW'15)*. Int. World Wide Web Conf.s Steering Committee, Geneva, Switzerland, 776–786. https://doi.org/10.1145/2736277.2741124

[90] Gaurav Misra and Jose M. Such. 2016. How Socially Aware Are Social Media Privacy Controls? *Computer* 49, 3 (Mar 2016), 96–99. https://doi.org/10.1109/MC.2016.83

[91] Miljana Mladenović, Vera Ošmjanski, and Staša Vujičić Stanković. 2021. Cyber-Aggression, Cyberbullying, and Cyber-Grooming: A Survey and Research Challenges. *ACM Comp. Surv.* 54, 1 (Jan. 2021), 1–42. https://doi.org/10.1145/3424246

[92] Francesca Mosca and Jose M Such. 2021. ELVIRA: an Explainable Agent for Value and Utility-driven Multiuser Privacy. In *20th Int. Conf. on Autonomous Agents and Multiagent Sys. (AAMAS'21)*. Int. Foundation for Autonomous Agents and Multiagent Sys., London, UK (virtual), In press.

[93] Francesca Mosca, Jose M Such, and Peter Mcburney. 2019. Value-driven Collaborative Privacy Decision Making. In *Proc. of the AAAI Spring Symp. on Priv.-Enhancing Artificial Intelligence and Language Tech. (PAL'19)*. CEUR Workshop Proc., Stanford, California, USA, 13–20. http://ceur-ws.org/Vol-2335/1st_PAL_paper_4.pdf

[94] Francesca Mosca, Jose M. Such, and Peter McBurney. 2020. Towards a Value-driven Explainable Agent for Collective Privacy. In *Proc. of the 19th Int. Conf. on Autonomous Agents and MultiAgent Sys. (AAMAS'20)*. Int. Foundation for Autonomous Agents and Multiagent Sys., Auckland, New Zealand, 1937–1939. https://doi.org/10.5555/3398761.3399033

[95] Daniel S Nagin. 1998. Criminal deterrence research at the outset of the twenty-first century. *Crime and Justice* 23 (1998), 1–42. https://doi.org/10.1086/449268

[96] Helen Nissenbaum. 2009. *Privacy in context: Technology, policy, and the integrity of social life.* Stanford University Press, Stanford, CA, USA. https://www.sup.org/books/title/?id=8862

[97] Patricia A Norberg, Daniel R Horne, and David A Horne. 2007. The privacy paradox: Personal information disclosure intentions versus behaviors. *Jour. of Consumer Affairs* 41, 1 (2007), 100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x

[98] State of California Department of Justice. 2018. California Consumer Privacy Act (CCPA). https://oag.ca.gov/privacy/ccpa Last accessed 2nd of October 2020.

[99] Council of Europe Portal. 1950. Convention for the Protection of Human Rights and Fundamental Freedoms. https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/0900001680063765 Last accessed 2nd of October 2020.

[100] Jason A. Okonofua, David Paunesku, and Gregory M. Walton. 2016. Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proc. of the National Academy of Sciences (PNAS)* 113, 19 (May 2016), 5221–5226. https://doi.org/10.1073/pnas.1523698113

[101] Alexandra-Mihaela Olteanu, Kévin Huguenin, Italo Dacosta, and Jean-Pierre Hubaux. 2018. Consensual and Privacy-Preserving Sharing of Multi-Subject and Interdependent Data. In *Proc. of the Symp. on Network and Distributed Sys. Secu. (NDSS'18)*. Internet Society, San Diego, CA, USA, 15. https://doi.org/10.14722/ndss.2018.23002

[102] José Ramón Padilla-López, Alexandros Andre Chaaraoui, and Francisco Flórez-Revuelta. 2015. Visual privacy protection methods: A survey. *Expert Sys. with Applications* 42, 9 (June 2015), 4177–4195. https://doi.org/10.1016/j.eswa.2015.01.041

[103] Leysia Palen and Paul Dourish. 2003. Unpacking" privacy" for a networked world. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'03)*. ACM, New York, NY, USA, 129–136. https://doi.org/10.1145/642611.642635

[104] Raymond Paternoster and Sally Simpson. 1996. Sanction threats and appeals to morality: Testing a rational choice model of corporate crime. *Law and Society Rev.* 30, 3 (1996), 549–583. https://doi.org/10.2307/3054128

[105] Sameer Patil, Roberto Hoyle, Roman Schlegel, Apu Kapadia, and Adam J. Lee. 2015. Interrupt Now or Inform Later? Comparing Immediate and Delayed Privacy Feedback. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'15)*. Assoc. for Comp. Mach. (ACM), New York, NY, USA, 1415–1418. https://doi.org/10.1145/2702123.2702165

[106] Sameer Patil, Roman Schlegel, Apu Kapadia, and Adam J. Lee. 2014. Reflection or Action? How Feedback and Control Affect Location Sharing Decisions. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'14)*. Assoc. for Comp. Mach. (ACM), New York, NY, USA, 101–110. https://doi.org/10.1145/2556288.2557121

[107] Sandra Petronio. 2010. Communication privacy management theory: What do we know about family privacy regulation? *Jour. of Family Theory & Rev.* 2, 3 (2010), 175–196. https://doi.org/10.1111/j.1756-2589.2010.00052.x

[108] Chanda Phelan, Cliff Lampe, and Paul Resnick. 2016. It's creepy, but it doesn't bother me. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'16)*. ACM, New York, NY, USA, 5240–5251. https://doi.org/10.1145/2858036.2858381

[109] José C Pinheiro and Douglas M Bates. 2000. *Mixed-effects models in S and S-PLUS*. Springer, New York, NY, USA. https://doi.org/10.1007/b98882

[110] Canada.ca Portal. 2019. Consolidated federal laws of canada, Protecting Canadians from Online Crime Act. https://laws-lois.justice.gc.ca/eng/annualstatutes/2014_31/page-1.html Last accessed 2nd of October 2020.

[111] EUR-Lex Portal. 2016. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. https://eur-lex.europa.eu/eli/reg/2016/679/oj Last accessed 2nd of October 2020.

[112] United Nations Portal. 1948. Universal Declaration of Human Rights. https://www.un.org/en/universal-declaration-human-rights/ Last accessed 2nd of October 2020.

[113] Felix Portnoy and Gary Marchionini. 2010. Modeling the effect of habituation on banner blindness as a function of repetition and search type: gap analysis for future work. In *Proc. of the ACM Conf. Extended Abstracts on Human Factors in Comp. Sys. (CHI EA'10)*. ACM, Atlanta, GA, USA, 4297–4302. https://doi.org/10.1145/1753846.1754142

[114] William L Prosser. 1960. Privacy. *Cali. Law Rev.* 48, 3 (Aug. 1960), 383–423. https://web.archive.org/web/20131019050717/http://www.californialawreview.org/assets/pdfs/misc/prosser_privacy.pdf

[115] Sarah Rajtmajer, Anna Cinzia Squicciarini, Jose M. Such, Justin Semonsen, and Andrew Belmonte. 2017. An Ultimatum Game Model for the Evolution of Privacy in Jointly Managed Content. In *Proc. of the Int. Conf. on Decision and Game Theory for Secu. (GameSec'17)*. Springer, Vienna, Austria, 112–130. https://doi.org/10.1007/978-3-319-68711-7_7

[116] 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 *Proc. of the Symp. on Usable Priv. and Secu. (SOUPS'18)*. USENIX Assoc., USA, 143–157. https://www.usenix.org/conference/soups2018/presentation/rashidi

[117] Yasmeen Rashidi, Apu Kapadia, Christena Nippert-Eng, and Norman Makoto Su. 2020. "It's easier than causing confrontation": Sanctioning Strategies to Maintain Social Norms of Content Sharing and Privacy on Social Media. *Proc. of the ACM Jour.: Human-Comp. Interaction: Comp. Supported Cooperative Work and Social Comp.* 4, CSCW1 (May 2020), 23:1–23:25. https://doi.org/10.1145/3392827

[118] Arunee Ratikan and Mikifumi Shikida. 2014. Privacy Protection Based Privacy Conflict Detection and Solution in Online Social Networks. In *Proc. of the Int. Conf. on Human Aspects of Info. Secu., Priv., and Trust (HAS'14)*. Springer, Heraklion, Crete, Greece, 433–445. https://doi.org/10.1007/978-3-319-07620-1_38

[119] Elissa M. Redmiles, Sean Kross, and Michelle M. Mazurek. 2019. How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples. In *Proc. of the IEEE Symp. on Secu. and Priv. (S&P'19)*. IEEE, San Francisco, CA, USA, 1326–1343. https://doi.org/10.1109/SP.2019.00014

[120] Robert AC Ruiter, Charles Abraham, and Gerjo Kok. 2001. Scary warnings and rational precautions: A review of the psychology of fear appeals. *Psychology and Health* 16, 6 (2001), 613–630. https://doi.org/10.1080/08870440108405863

[121] Robert AC Ruiter, Loes TE Kessels, Gjalt-Jorn Y Peters, and Gerjo Kok. 2014. Sixty years of fear appeal research: Current state of the evidence. *Int. Jour. of Psychology* 49, 2 (2014), 63–70. https://doi.org/10.1002/ijop.12042

[122] Clayton Santos, Eulanda M. dos Santos, and Eduardo Souto. 2012. Nudity detection based on image zoning. In *2012 11th Int. Conf. on Info. Science, Signal Processing and their Applications (ISSPA'12)*. IEEE, New York, NY, USA, 1098–1103. https://doi.org/10.1109/ISSPA.2012.6310454

[123] Joseph Schmid and Sofia Bouderbala. 2018. Facebook denies 'censoring' 19th-century vagina painting. https://phys.org/news/2018-02-facebook-denies-censoring-19th-century-vagina.html Last accessed 6th of October 2020.

[124] Claudia R. Schneider, Lisa Zaval, Elke U. Weber, and Ezra M. Markowitz. 2017. The influence of anticipated pride and guilt on pro-environmental decision making. *PLOS ONE* 12, 11 (Nov. 2017), e0188781. https://doi.org/10.1371/journal.pone.0188781 Publisher: Public Library of Science.

[125] Amelia Schonbek. 2019. When Women Don't Want to Talk. https://www.thecut.com/2019/10/when-women-dont-want-to-talk-about-assault-and-harassment.html Last accessed 25th of May 2020.

[126] Ivor Shapiro and Brian MacLeod Rogers. 2017. How the "Right to be Forgotten" Challenges Journalistic Principles: Privacy, freedom and news durability. *Digital Journalism* 5, 9 (Oct 2017), 1101–1115. https://doi.org/10.1080/21670811.2016.1239545

[127] Lijiang Shen. 2010. Mitigating psychological reactance: The role of message-induced empathy in persuasion. *Human Comm. Research* 36, 3 (2010), 397–422. https://doi.org/10.1111/j.1468-2958.2010.01381.x

[128] Lijiang Shen. 2011. The effectiveness of empathy-versus fear-arousing antismoking PSAs. *Health Comm.* 26, 5 (2011), 404–415. https://doi.org/10.1080/10410236.2011.552480

[129] Anna Squicciarini, Cornelia Caragea, and Rahul Balakavi. 2017. Toward Automated Online Photo Privacy. *ACM Trans. Web* 11, 1, Article 2 (April 2017), 29 pages. https://doi.org/10.1145/2983644

[130] Anna Cinzia Squicciarini, Mohamed Shehab, and Federica Paci. 2009. Collective privacy management in social networks. In *Proc. of the ACM Int. Conf. on World Wide Web (WWW'09)*. ACM, Madrid, Spain, 521–530. https://doi.org/10.1145/1526709.1526780

[131] Anna Cinzia Squicciarini, Mohamed Shehab, and Joshua Wede. 2010. Privacy policies for shared content in social network sites. *The VLDB Jour.* 19, 6 (Dec 2010), 777–796. https://doi.org/10.1007/s00778-010-0193-7

[132] Keith Stanovich. 2010. *Rationality and the Reflective Mind.* Oxford University Press, New York, NY, USA. 319 pages. https://doi.org/10.1093/acprof:oso/9780195341140.001.0001

[133] Fritz Strack and Roland Deutsch. 2004. Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Rev.* 8, 3 (aug 2004), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

[134] Scott R. Stroud. 2014. The Dark Side of the Online Self: A Pragmatist Critique of the Growing Plague of Revenge Porn. *Jour. of Mass Media Ethics* 29, 3 (Jul 2014), 168–183. https://doi.org/10.1080/08900523.2014.917976

[135] Jose M. Such and Natalia Criado. 2014. Adaptive Conflict Resolution Mechanism for Multi-party Privacy Management in Social Media. In *Proc. of the ACM Workshop on Priv. in the Electronic Society (WPES'14)*. ACM, Scottsdale, AZ, USA, 69–72. https://doi.org/10.1145/2665943.2665964

[136] Jose M. Such and Natalia Criado. 2016. Resolving Multi-Party Privacy Conflicts in Social Media. *IEEE Trans. on Knowledge and Data Engineering* 28, 7 (Jul 2016), 1851–1863. https://doi.org/10.1109/TKDE.2016.2539165

[137] Jose M. Such and Natalia Criado. 2018. Multiparty privacy in social media. *Comm. of the ACM* 61, 8 (Jul 2018), 74–81. https://doi.org/10.1145/3208039

[138] Jose M. Such, Joel Porter, Sören Preibusch, and Adam Joinson. 2017. Photo Privacy Conflicts in Social Media: A Large-scale Empirical Study. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'17)*. ACM, Denver, CO, USA, 3821–3832. https://doi.org/10.1145/3025453.3025668

[139] Jose M. Such and Michael Rovatsos. 2016. Privacy Policy Negotiation in Social Media. *ACM Trans. on Autonomous and Adaptive Sys. (TAAS)* 11, 1 (Feb 2016), 1–29. https://doi.org/10.1145/2821512

[140] Richard H. Thaler and Cass R. Sunstein. 2008. *Nudge: improving decisions about health, wealth, and happiness.* Yale University Press, New Haven, CT, USA. https://en.wikipedia.org/wiki/Nudge_(book)

[141] Kurt Thomas, Chris Grier, and David M. Nicol. 2010. unFriendly: Multi-party Privacy Risks in Social Networks. In *Priv. Enhancing Tech.* Springer, Berlin, Germany, 236–252. https://doi.org/10.1007/978-3-642-14527-8_14

[142] Khai N. Truong, Gillian R. Hayes, and Gregory D. Abowd. 2006. Storyboarding: an empirical determination of best practices and effective guidelines. In *Proc. of the ACM Conf. on Designing Interactive Sys. (DIS)*. ACM, University Park, PA, USA, 12. https://doi.org/10.1145/1142405.1142410

[143] 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. https://doi.org/10.1177/1461444814543164

[144] Anthony Vance, Brock Kirwan, Daniel Bjornn, Jeffrey Jenkins, and Bonnie Brinton Anderson. 2017. What Do We Really Know about How Habituation to Warnings Occurs Over Time? A Longitudinal FMRI Study of Habituation and Polymorphic Warnings. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'17)*. ACM, New York, NY, USA, 2215–2227. https://doi.org/10.1145/3025453.3025896

[145] Nishant Vishwamitra, Yifang Li, Kevin Wang, Hongxin Hu, Kelly Caine, and Gail-Joon Ahn. 2017. Towards PII-based Multiparty Access Control for Photo Sharing in Online Social Networks. In *Proc. of the ACM on Symp. on*

Access Control Models and Tech. (SACMAT'17). ACM, Indianapolis, IN, USA, 155–166. https://doi.org/10.1145/3078861.3078875

[146] Kate Walker and Emma Sleath. 2017. A systematic review of the current knowledge regarding revenge pornography and non-consensual sharing of sexually explicit media. *Aggression and Violent Behavior* 36 (Sep 2017), 9–24. https://doi.org/10.1016/j.avb.2017.06.010

[147] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A field trial of privacy nudges for facebook. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'14)*. ACM, Toronto, ON, Canada, 2367–2376. https://doi.org/10.1145/2556288.2557413

[148] 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 *Proc. of the 22nd Int. Conf. on World Wide Web (WWW'13)*. ACM, Rio de Janeiro, Brazil, 763–770. https://doi.org/10.1145/2487788.2488038

[149] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011. "I regretted the minute I pressed share": a qualitative study of regrets on Facebook. In *Proc. of the Symp. on Usable Priv. and Secu. (SOUPS'11)*. ACM, Pittsburgh, PA, USA, 1–16. https://doi.org/10.1145/2078827.2078841

[150] Ryan Wishart, Domenico Corapi, Srdjan Marinovic, and Morris Sloman. 2010. Collaborative Privacy Policy Authoring in a Social Networking Context. In *Proc. of the IEEE Int. Symp. on Policies for Distributed Sys. and Networks (POLICY'10)*. IEEE, Fairfax, VA, USA, 1–8. https://doi.org/10.1109/POLICY.2010.13

[151] Pamela Wisniewski, AKM Islam, Heather Richter Lipford, and David C Wilson. 2016. Framing and measuring multidimensional interpersonal privacy preferences of social networking site users. *Comm. of the Assoc. for Info. Sys.* 38, 1 (2016), 235–258. https://doi.org/10.17705/1CAIS.03810

[152] Pamela Wisniewski, Heather Lipford, and David Wilson. 2012. Fighting for my space: coping mechanisms for SNS boundary regulation. In *Proc. of the ACM Conf. on Human Factors in Comp. Sys. (CHI'12)*. ACM, Austin, TX, USA, 609–618. https://doi.org/10.1145/2207676.2207761

[153] Pamela Wisniewski, Heng Xu, Heather Lipford, and Emmanuel Bello-Ogunu. 2015. Facebook apps and tagging: The trade-off between personal privacy and engaging with friends. *Jour. of the Assoc. for Info. Science and Tech.* 66, 9 (2015), 1883–1896. https://doi.org/10.1002/asi.23299

[154] Kim Witte and Mike Allen. 2000. A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health Education & Behavior* 27, 5 (2000), 591–615. https://doi.org/10.1177/109019810002700506

[155] Sixuan Zhang, Liang Yu, Robin L. Wakefield, and Dorothy E. Leidner. 2016. Friend or Foe: Cyberbullying in Social Network Sites. *SIGMIS Database* 47, 1 (Feb. 2016), 51–71. https://doi.org/10.1145/2894216.2894220

[156] Haoti Zhong, Anna Squicciarini, and David Miller. 2018. Toward Automated Multiparty Privacy Conflict Detection. In *Proc. of the 27th ACM Int. Conf. on Info. and Knowledge Management (CIKM'18)*. Assoc. for Comp. Mach. (ACM), New York, NY, USA, 1811–1814. https://doi.org/10.1145/3269206.3269329

Received June 2020; revised October 2020; accepted December 2020