New research on Apple's child safety feature shows warnings can increase risky sharing

September 28 2021, by Bennett Bertenthal, Apu Kapadia, Kurt Hugenberg

Credit: Unsplash/CC0 Public Domain

Apple's plan to roll out tools to limit the spread of child sexual abuse material has drawn praise from some privacy and security experts as well

as by child protection advocacy groups. There has also been an outcry about invasions of privacy.

These concerns have obscured another even more troublesome problem that has received very little attention: Apple's new feature uses design elements shown by research to backfire.

One of these new features adds a parental control option to Messages that blocks the viewing of sexually explicit pictures. The expectation is that parental surveillance of the child's behavior will decrease the viewing or sending of sexually explicit photos, but this is highly debatable.

We are two psychologists and a computer scientist. We have conducted extensive research on why people share risky images online. Our recent research reveals that warnings about privacy on social media do not reduce photo-sharing nor increase concern about privacy. In fact, these warnings, including Apple's new child safety features, can increase rather than reduce risky sharing of photos.

Apple's child safety features

Apple announced on Aug. 5, 2021 that it plans to introduce new child safety features in three areas. The first, relatively uncontroversial feature is that Apple's search app and virtual assistant Siri will provide parents and children with resources and help if they encounter potentially harmful material.

The second feature will scan images on people's devices that are also stored in iCloud Photos to look for matches in a database of child sexual abuse images provided by the National Center for Missing and Exploited Children and other child safety organizations. After a threshold for these matches is reached, Apple manually reviews each machine match to

confirm the content of the photo, and then disables the user's account and sends a report to the center. This feature has generated much controversy.

The last feature adds a parental control option to Messages, Apple's texting app, that blurs sexually explicit pictures when children attempt to view them. It also warns the children about the content, presents helpful resources and assures them it is OK if they do not want to view the photo. If the child is 12 or under, parents will get a message if the child views or shares a risky photo.

There has been little public discussion of this feature, perhaps because the conventional wisdom is that parental control is necessary and effective. This is not always the case, however, and such warnings can backfire.

When warnings backfire

In general, people are more likely than not to avoid risky sharing, but it's important to reduce the sharing that does occur. An analysis of 39 studies found that 12% of young people forwarded a sext, or sexually explicit image or video, without consent, and 8.4% had a sext of themselves forwarded without consent. Warnings might seem like an appropriate way to do so. Contrary to expectation, we have found that warnings about privacy violations often backfire.

In one series of experiments, we tried to decrease the likelihood of sharing embarrassing or degrading photos on social media by reminding participants that they should consider the privacy and security of others. Across multiple studies, we have tried different reminders about the consequences of sharing photos, similar to the warnings to be introduced in Apple's new child safety tools.

Remarkably, our research often reveals paradoxical effects. Participants who received warnings as simple as stating that they should take others' privacy into account were more likely to share photos than participants who did not receive this warning. When we began this research, we were sure that these privacy nudges would reduce risky photo sharing, but they didn't.

The results have been consistent since our first two studies showed that warnings backfired. We have now observed this effect multiple times, and have found that several factors, such as a person's humor style or photo sharing experience on social media, influence their willingness to share photos and how they might respond to warnings.

Although it's not clear why warnings backfire, one possibility is that individuals' concerns about privacy are lessened when they underestimate the risks of sharing. Another possibility is reactance, or the tendency for seemingly unnecessary rules or prompts to elicit the opposite effect from what was intended. Just as a forbidden fruit becomes sweeter, so too might constant reminders about privacy concerns make risky photo sharing more attractive.

Will Apple's warnings work?

It is possible that some children will be more inclined to send or receive sexually explicit photos after receiving a warning from Apple. There are numerous reasons why this behavior may occur, ranging from curiosity—adolescents often learn about sex from peers—to challenging parents' authority and reputational concerns, such as being seen as cool by sharing apparently risky photos. During a stage of life when risk-taking tends to peak, it's not hard to see how adolescents might find earning a warning from Apple to be a badge of honor rather than a genuine cause for concern.

Apple announced on Sept. 3, 2021 that it is delaying the rollout of these new CSAM tools because of concerns expressed by the privacy and security community. The company plans to take additional time over the coming months to collect input and make improvements before releasing these child safety features.

This plan is not sufficient, however, without also knowing whether Apple's new features will have the desired effect on children's behavior. We encourage Apple to engage with researchers to ensure that their new tools will reduce rather than encourage problematic photo sharing.

This article is republished from The Conversation under a Creative Commons license. Read the original article.

Provided by The Conversation

Citation: New research on Apple's child safety feature shows warnings can increase risky sharing (2021, September 28) retrieved 26 April 2024 from https://techxplore.com/news/2021-09-apple-child-safety-feature-risky.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.