UC BERKELEY

CENTER FOR LONG-TERM CYBERSECURITY

CLTC WHITE PAPER SERIES

Designing Risk Communications

A ROADMAP FOR DIGITAL PLATFORMS

JESSICA NEWMAN, ANN CLEAVELAND, GRACE GORDON, AND STEVEN WEBER

CLTC WHITE PAPER SERIES

Designing Risk Communications

A ROADMAP FOR DIGITAL PLATFORMS

JESSICA NEWMAN, ANN CLEAVELAND, GRACE GORDON, AND STEVEN WEBER

DECEMBER 2020

CENTER FOR LONG-TERM CYBERSECURITY

University of California, Berkeley

Contents

Executive Summary 1
Introduction 3

PART I: THE CHALLENGES OF RISK COMMUNICATION 6
What is Risk Communication? 6
Digital Risk Communication 7
- Communication Types 7
- Digital Risks 9
- Legal Requirements 10
- Sub-Legal Practices 12
- Examples of Digital Risk Communication 14
  - Privacy Nutrition Labels 14
  - The Privacy Checkup Tool 16
  - Transparency Reports 17
  - Safety Reports 18

Gaps, Complexities, and Tensions 19
- Transparency 19
- Distrust 20
- Scalability 20
- Uncertainty 21

PART II: THEORY AND CROSS-INDUSTRY LEARNING 22
The Science of Risk Communication 22
- Relevant Themes in Risk Communication Theory 23
- Critiques and Evolution of Thought 24

Lessons from Other Sectors 25
- Natural Disaster Response & Recovery: 2011 Mississippi River Flood 25
- Medical Products: Drug Facts Box 26
- Public Health: Communications about Zika Virus 28
- Food and Beverage Industry: Front-of-Pack Labeling 29
- Transportation: Aviation Safety Reporting System 30

PART III: A PRACTICAL AGENDA FORWARD **32**
Risk Communication Roadmap for Digital Platforms **32**
Engage **33**
Design **34**
Evaluate **35**

Conclusion **36**
About the Authors **38**
Acknowledgments **39**

Executive Summary

When digital platform companies become aware that their users have been exposed to risks, or are likely to be exposed in the future, they often do not have a playbook of effective communication practices that go beyond narrow notification requirements. The standard approaches — ranging from lengthy legal documents to easily dismissed pop-ups — typically fail to inform users in ways that enable better decision-making.

This paper looks at some existing practices used by digital platforms, and identifies common challenges and tensions. We define *digital platforms* here as online products and services that facilitate interactions between at least two different groups.[1] This includes social media companies, such as Facebook or Twitter, but also extends to services as varied as ride-sharing (e.g. Uber), healthcare (e.g. glucose-monitoring apps), and consumer internet of things (IoT) devices (e.g. digital personal assistants). We focus on risks that fall below legal reporting thresholds, as this is where firms have the most freedom of action. This is also where firms have the most to gain or lose in terms of user trust and reputation.

The paper explores relevant insights from theories of risk communications, how these have been applied in practice across different sectors, and why they are relevant to the particular risks and communication needs of digital platforms. The analysis was informed by a series of expert interviews and a multi-stakeholder workshop held in February 2020 at the University of California, Berkeley to investigate the current state of risk communications in the technology sector, and to assess the feasibility of adopting best practices from other sectors and the scientific literature. Finally, we propose a practical agenda, introducing a "roadmap" to serve as a preliminary guide for digital platforms trying to redefine how they communicate risks with users. The roadmap draws from reliable insights at the intersection of decision science, psychology, sociology, and communications to propose three practices that digital platform firms can adopt to improve risk communications, including:

Engagement: *Build trusting relationships with users based on transparent, comprehensive dialogue to facilitate effective risk communication.*

1 Information Technology & Innovation Foundation, "ITIF Technology Explainer: What Are Digital Platforms?" October 12, 2018, https://itif.org/publications/2018/10/12/itif-technology-explainer-what-are-digital-platforms.

Design: *Create accessible, informative, and actionable communication formats to enable effective risk communication.*

Evaluation: *Establish processes for risk communication in advance, and create metrics to assess effectiveness and ensure the resilience of risk communication efforts.*

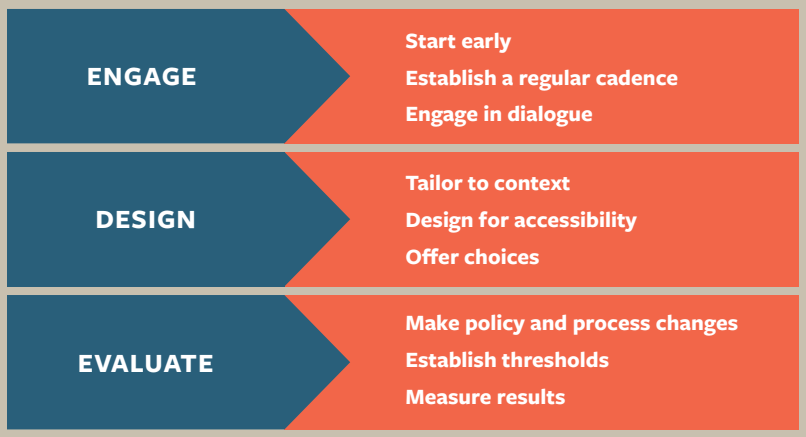


Figure I. Overview of the Risk Communication Roadmap for Digital Platforms

Introduction

The proliferation of digital technologies not only has altered and amplified people's exposure to certain risks, including security threats and privacy violations, but also has reshaped how information about such risks is communicated. Amidst the cacophony of voices on the internet, communications about relevant risks can be hard to decipher. This reality has exposed a weakness of digital platforms: although they are skilled at enabling communication in general, they are less adept at ensuring the effectiveness of critical communications. Companies that become aware of risks their users have been exposed to, or are likely to be exposed to in the future, typically do not have a playbook of effective communication practices that goes beyond narrow notification requirements. As a concrete example, imagine that a company discovers that sensitive data about a small group of users was exposed in an unencrypted database for a few hours and may have been accessed. If this event falls below the legal threshold for required notification, what (if anything) should the company say, and to whom?

Communicating about risks has never been easy, and many industries have grappled with this challenge. There is also a rich body of literature about the science and theory of risk communications that spans decades, as well as lessons that have been learned over time. However, these insights have not been reliably translated to the domain of digital platforms, despite the urgency and scale of risks at stake, which span security, privacy, and other potential harms. It has become axiomatic that people have the right to know how their information is collected and processed as they engage with digital services. However, current communication methods, such as inscrutable legal disclosures and pop-up alerts, too often fail to meaningfully inform people because they are rarely accessible, relatable, or actionable. Privacy policies, for example, are notoriously difficult and time-consuming to interpret and are widely ignored by users.[2] Despite efforts from regulators and companies to facilitate more user-friendly notification practices, firms have been inconsistent in their adoption of improved practices.[3]

The relatively narrow frames of security and privacy notifications, which are the focus of the majority of the relevant legislation around the world, have not yielded sufficient results in the service of users' interests. They also have left significant gray areas for companies regarding

2 Jonathan A. Obar & Anne Oeldorf-Hirsch, "The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services," *Information, Communication & Society*, 23:1, 128-147, 2020, https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1486870.

3 Erin Egan, "Communicating About Privacy: Towards People-Centered and Accountable Design," Facebook, July 2020, https://about.fb.com/wp-content/uploads/2020/07/Privacy-Transparency-White-Paper.pdf.

incidents that fall below legal notification thresholds. This paper suggests that a broader framework of risk communication can help companies overcome some of these hurdles, and can benefit users and firms at the same time. Decades of literature about the science of risk communication, combined with lessons from other sectors, provide a fresh take on how digital platforms can improve meaningful dialogue and engagement with users about potential risks.

Risk communication is a challenge for organizations and institutions of all kinds, but in this paper, we focus on the specific challenge of risk communications between digital platforms and users. In particular, we focus on security and privacy risks, including some of the most common ways in which data can be exposed or exploited. We define digital platforms as online products and services that facilitate interactions between at least two different groups.[4] This includes social media companies, such as Facebook or Twitter, but also extends to services as varied as ride-sharing (e.g. Uber), healthcare (e.g. a glucose-monitoring app), and consumer IoT devices (e.g. digital personal assistants). A failure to effectively communicate risks can have harmful impacts on users, and platforms are increasingly acknowledging the need for changes to facilitate effective, user-centered communication.[5]

Many digital platforms already face a trust deficit with their users (and with regulators and other stakeholders), and so need to ensure that their risk communications are not perceived as a form of manipulation or a "Band-Aid" applied to avoid dealing with underlying problems. Though largely beyond the scope of this paper, firms must continue efforts to address the structural weaknesses that generate security and privacy risks to users, whether through technical, organizational, or governance mechanisms. However, privacy and security will likely never be completely "fixed," and there will always be vulnerabilities. It is therefore critical that firms have a communication plan in place, and, as this paper describes, there are better and worse ways to do that.

We start from the proposition that digital platforms have a responsibility to protect the data over which they have custody or control, and to be transparent when users' data may be exposed or exploited.[6] But these terms are underspecified in practice: how much responsibility do firms have, for what risks are they responsible, and what level of transparency is

4 Information Technology & Innovation Foundation, "ITIF Technology Explainer: What Are Digital Platforms?"

5 Erin Egan, "Communicating About Privacy: Towards People-Centered and Accountable Design," Facebook, July 2020, https://about.fb.com/wp-content/uploads/2020/07/Privacy-Transparency-White-Paper.pdf.

6 Organization for Economic Cooperation and Development, "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data," Guidelines 12 & 13 (originally published 1980, updated 2013), https://www.oecd.org/sti/ ieconomy/oecd_privacy_framework.pdf.

high enough? As expectations evolve, firms have an opportunity to shape norms and practices. Because many people rely on multiple digital platforms at once (for example, using a social media platform, a ride-sharing app, and a dating site all on any given day), the practices of one company are likely to affect the playing field for others. Since platforms face many common risks, when one platform raises the bar, people may come to expect the same from others. For example, many security and privacy experts expect that Apple's new "privacy nutrition labels" — through which developers must display what data their apps collect — will lead to improved practices across other platforms.[7] Companies that demonstrate the best risk communication practices have an opportunity to establish *de facto* standards across the industry. Conversely, companies that fail to communicate risks effectively may foster distrust. This interdependency suggests that shared frameworks for risk communications may be helpful for the industry as a whole.

This paper examines existing practices used by digital platforms and identifies pressing gaps, complexities, and tensions. We focus on risks that fall below legal reporting thresholds because this is where firms have the most freedom to experiment. This is also the area where firms have the most to gain or lose in terms of user trust and reputation. The paper then explores the theory of risk communications, how it has been applied across different sectors, and why it is relevant to the particular needs of digital platforms. Finally, the paper puts forward elements of a practical agenda, introducing a roadmap that provides a preliminary guide for digital platforms trying to redefine the role of risk communication in their relationships with users.

7 Maria Henriquez, "Apple's new requirement puts additional focus on consumer and data privacy," *Security*, November 11, 2020, https://www.securitymagazine.com/articles/93903-apples-new-requirement-puts-consumer-and-data-privacy-at-the-forefront.

Part I: The Challenges of Risk Communication

WHAT IS RISK COMMUNICATION?

A widely accepted and pragmatic definition describes risk communication as a process for explaining information people need in a manner they can use to make informed decisions about risks to their health, safety, and environment.[8] The goals of risk communication vary among different stakeholders. Users may rely on risk communications to understand the risks, costs, and benefits of using a service,[9] while institutions and firms often see risk communication as a way to help "minimize disputes, resolve issues, and anticipate problems before they result in an irreversible breakdown in communication," with the ultimate objective of preventing irreversible damage to the business.[10] Firms have struggled to develop risk communications that are not only adequate and compliant for their legal teams, but demonstrably beneficial for the users and communities they serve. Understanding risk communications as the simple "transfer" or exchange of usable information also obscures that effective communication is multi-directional, and that risk communication should be an ongoing process of engagement among firms, customers, and communities. This is especially true when risks are uncertain, when they change rapidly, or when they generate disparate impacts for different people and groups. Rather than a series of one-way interactions, risk communication needs to be an iterated, multi-party process that evolves over time.

8 M. Granger Morgan, Baruch Fischhoff, Ann Bostrom, Cynthia J. Atman. *Risk Communication: A Mental Models Approach*. (Cambridge University Press, 2002.) While risk communication can be an important component of crisis communications, for this analysis, we distinguish *risk communication* (which supports deliberative decisions and actions) from *crisis communication* (which requires immediate response, often under conditions of stress).

9 National Research Council (U.S.), editor. *Improving Risk Communication*. National Academy Press, 1989.

10 Ragnar Löfstedt, "Risk Communication and Management in the 21st Century," Regulation2point0, Working Papers, (April 2004). Available at SSRN: https://ssrn.com/abstract=545724.

DIGITAL RISK COMMUNICATION

Risk communications for digital platforms present a specific case of this general problem. This section provides background and lays out the current landscape of risk communication for digital platforms, including a discussion of common digital risks, risk categorization schemes, and communication types. It then covers both the legal requirements and sub-legal practices seen in digital risk communication, and identifies common gaps, complexities, and tensions that emerge.

Communication Types

Lorrie Faith Cranor (Director of Carnegie Mellon University's CyLab Usable Privacy and Security Laboratory) identifies five types of communication that are relevant to digital security tasks:[11]

1. **Warnings.** Warnings alert users about a hazard, and sometimes also include immediate actions that can be taken to avoid the hazard. They often take the form of pop-up alerts or warning indicators.
2. **Notices.** Notices provide information about the characteristics of an entity or object. For example, privacy policies and SSL certificates are common types of notices.
3. **Status Indicators.** Status indicators inform users about system status information — for example, menu bar indicators that show whether anti-virus software is up to date.
4. **Training.** Training communications are intended to teach users about security threats and how to respond to them. They can take many forms, including tutorials, games, web sites, emails, courses, and videos.
5. **Policies.** Policy communications inform users about relevant policies they are expected to comply with — for example, organizational policies around passwords. Policies may be communicated via terms of service documents or memos.

Cranor further differentiates security communications along a design scale from passive to active. She describes active communications as "designed to interrupt the user's primary task and force them to pay attention," while passive communications are described as "available to the user, but easily ignored." For example, an active communication may prevent you from

11 Lorrie Faith Cranor, "A Framework for Reasoning About the Human in the Loop," UPSEC'08: Proceedings of the 1st Conference on Usability, Psychology, and Security, April 2008, https://dl.acm.org/doi/10.5555/1387649.1387650.

loading a website that appears unsafe without explicit intention, while a passive communication may simply involve a change in the color of an icon that signals a warning, but nothing further. Different risks — and their associated severity, scale, and frequency, as well as whether they can be mitigated through user actions — all call for different communication formats that user interface designers must parse. Passive communication methods are often more appropriate if risks are frequent, low severity, and non-actionable. Conversely, interface designers often choose an active communication method to alert users about risks that are rare or dangerous, or that can be quickly managed through an action.

Risk communication is most effective when it is tied to actions that can be taken to manage risk, rather than simply inform about risk. However, there are many obstacles between communication and behavior change that complicate this objective. Cranor's human-in-the-loop security framework (*Figure II*) highlights several impediments that can get in the way of a user properly receiving an intended communication. As the framework suggests, behavior change requires more than just receiving a communication. It also requires processing and applying the knowledge. The difficulties of translating communication into behavior change are well documented for digital privacy and security risks. As information law expert Natali Helberger puts it, "informing consumers does not automatically lead to informed consumers."[12]

The potential ineffectiveness of risk communication for digital platforms motivates the work explored in this paper, and the roadmap described below includes both design and evaluation as key components of risk communication. Without ongoing assessment, there is a very real danger that communication will work to create a false sense of security and trust, an illusion of consumer empowerment, when it may in fact be generating confusion or frustration.[13] Helberger emphasizes that effective communication with users therefore cannot be a one-time act, but rather is a process that includes ongoing research on the behaviors and needs of users at various stages of information processing.[14] Helberger argues that effective communications have various qualities, including that they are well timed, standardized and machine-readable, written for users and not for lawyers, and framed in a way that responds to people's actual needs. None of these qualities can be defined abstractly and permanently; they need to be empirically assessed for effectiveness over time.

12 Natali Helberger, "Forms Matter: Informing consumers effectively," Study commissioned by BEUC, September 2013, https://www.ivir.nl/publicaties/download/Form_matters.pdf.

13 *Ibid.*

14 *Ibid.*

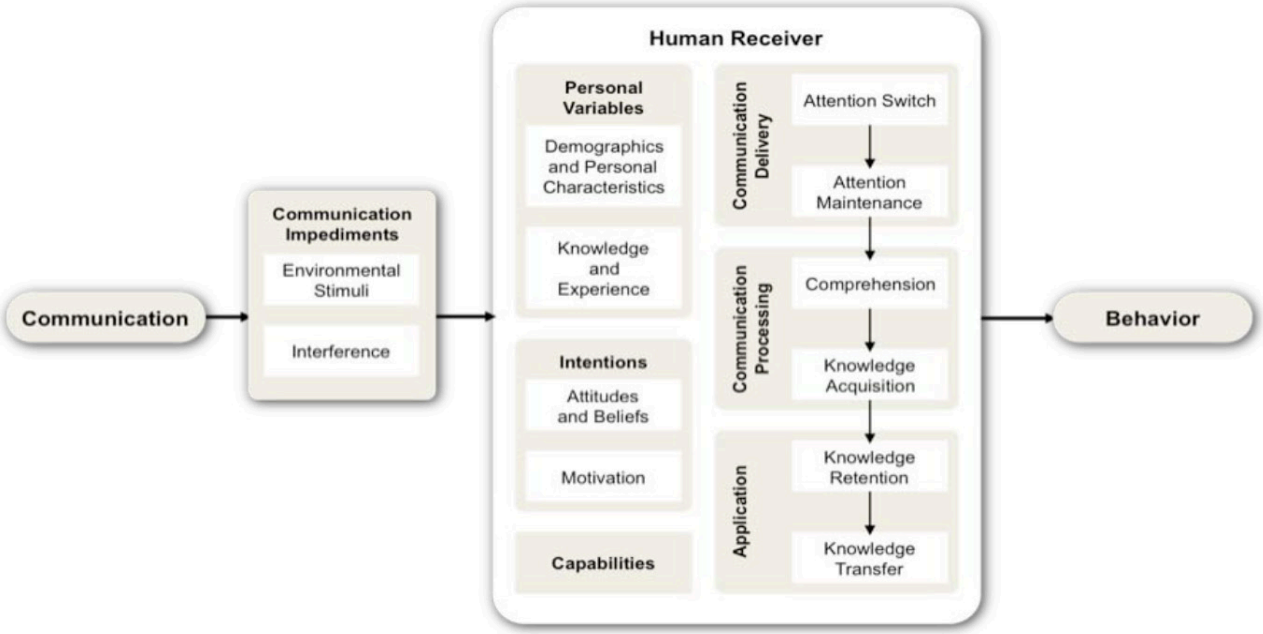


Figure II. The human-in-the-loop security framework illustrates the complex mix of variables that can obstruct translation of a communication into behavior change.[15]

Digital Risks

This paper is primarily concerned with managing communications about the risks of malicious or unauthorized exploitation of user data collected by digital platforms. This includes some privacy risks (such as possible access to personal information), and many security risks, meaning threats to the confidentiality, integrity, or availability of information or a service. We are particularly interested in those cases that fall below existing legal thresholds (for example, comparatively small data breaches, persistent threats such as phishing campaigns, or cases where there is uncertainty about the degree of harm), where standard communication practices do not yet exist. Major data breaches typically receive immediate attention, and firms have explicit communication guidelines and existing legal frameworks to manage such incidents. But these (more common) sub-legal threshold incidents, which may only receive attention following a delay, if at all, tend to have the most pervasive long-term impacts on a brand's reputation.[16] How a firm responds to these more minor risks sends an informative signal to customers, regulators, and

15 Lorrie Faith Cranor, "A Framework for Reasoning About the Human in the Loop," UPSEC'08: Proceedings of the 1st Conference on Usability, Psychology, and Security, April 2008, https://dl.acm.org/doi/10.5555/1387649.1387650.

16 Melanie Ensign, "What is a Security or Privacy Incident? Hiccups, F*ck Ups, and Give Ups," Discernible, July 2, 2020, https://medium.com/discernible/security-privacy-incident-hiccups-f-ck-ups-and-give-ups-e972ef46c3d.

the public about the firm's values and trustworthiness, precisely because whatever actions are taken in these moments is entirely at the firm's discretion.[17]

The increased usage of certain technologies has expanded the scale and scope of digital risks in recent years. For example, the implementation of artificial intelligence (AI) and machine learning technologies throughout many digital services has complicated calls for transparency in data collection and processing due to the difficulty of explaining how an AI model came to a particular decision. This can make it difficult for companies to live up to their stated goals of "transparency" about use of consumer data. Virtual reality and augmented reality also pose novel security and privacy risks to users because new types of sensitive data may be collected (for example, about a person's physical location and characteristics) and traditional communication interfaces, such as pop-up notifications, might not be feasible.[18]

Some potential risks to users of digital platforms, such as psychological and cognitive impacts related to screen addiction, manipulation, online harassment, and cyber bullying, are not explicitly addressed here and do not feature prominently in the existing literature, but such risks may become more important for digital platforms to address over time. Our focus on "users" also puts out of scope potential risks to employees and bystanders, as well as broader societal risks related to harms such as misinformation and anti-competitive behaviors that impact the economy, politics, and the environment. These risks are real and would benefit from improved communication strategies, but also demand more comprehensive responses and reevaluations that are beyond the scope of this paper.

Legal Requirements

Only a narrow subset of risks associated with digital platforms has explicit communication guidelines in existing legal frameworks. For example, companies are subject to data breach and security incident notification and disclosure laws across state, national, and transnational jurisdictions. Even for large firms, navigating the granular regulatory landscape can be challenging.

17 The classic explanation of this "costly signaling" argument comes from Thomas C. Schelling, *The Strategy of Conflict*, Harvard University Press, 1960.

18 James Pierce, Richmond Y. Wong, and Nick Merrill. 2020. Sensor Illumination: Exploring Design Qualities and Ethical Implications of Smart Cameras and Image/Video Analytics. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–19. DOI:https://doi.org/10.1145/3313831.3376347

While there is no overarching U.S. federal data breach notification law, many states have developed their own laws, which vary based on their treatment of such details as time constraints and definitions of "personal information." These existing legal requirements provide a rough baseline for the kinds of communications users can expect when something has gone wrong due to mishandling or misuse of their data, but do not provide consistent or clear direction across sectors or geographies. Arguments about the relative effectiveness of these laws are beyond the scope of this paper, but several landmark examples provide context about current practices and norms.

For example, if an unauthorized person has acquired a California resident's unencrypted personal information, the company responsible for holding that information is required under California law to notify the individual.[19] This notification must be delivered as soon as reasonably possible, and the type of communication required is explicitly defined: it must be titled "Notice of Data Breach," written in plain language in larger than 10-point type, and it must include headings detailing "What Happened," "What Information Was Involved," "What We Are Doing," "What You Can Do," and "For More Information."[20] The notification must also include a general description of the incident and a list of the types of personal information exposed during the breach, among other requirements.

The California Consumer Privacy Act (CCPA), which went into effect on January 1, 2020, extended these requirements.[21] CCPA gives consumers the right to request information about their personal information held by businesses, including how it is used and with whom it is shared, and they may ask for it to be deleted. CCPA also requires companies to give consumers certain notices explaining their privacy practices. These notices must list the categories of personal information collected and the purposes for which they are used. The notices must also link back to the company's full privacy policy.

There are also numerous federal laws that govern sector-specific notification requirements for businesses. For example, the HIPAA Breach Notification Rule issued by the U.S. Department of Health and Human Services, and the Health Breach Notification Rule issued by the Federal Trade Commission (FTC), require notification for breaches of protected health information and personal health records. The laws require that affected individuals are personally informed without unreasonable delay and provided with a description of what happened and how the

19 State of California Department of Justice, "Data Security Breach Reporting," https://oag.ca.gov/privacy/databreach/reporting.

20 See State of California Civil Code Section 1798.82 [amended by Stats. 2019, Ch. 750, Sec. 3. (AB 1130)], effective January 1, 2020. https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CIV&sectionNum=1798.82.

21 California Consumer Privacy Act of 2018, https://oag.ca.gov/privacy/ccpa.

breach is being investigated, what information was involved, how they can protect themselves from potential harm, and what the covered entity is doing to mitigate the harm and prevent further breaches. The U.S. Securities and Exchange Commission (SEC) also provides interpretative guidance on how companies should disclose security and privacy incidents; though this is directed primarily at investors, it further complicates the legal landscape for many firms.

Internationally, the landmark regulation for notification obligations is the European Union's General Data Protection Regulation (GDPR), which went into effect in May 2018. The GDPR requires companies to inform affected individuals about personal data breaches when there is a significant risk to their rights and freedoms. Companies are also required to inform the supervisory authority in the country or countries with affected users within 72 hours of knowledge of a data breach event, including information such as the nature and scope of the breach, its consequences, and the measures taken to mitigate its effects. The 72-hour breach notification requirement was a new requirement for the European Union and significantly changed reporting practices for many companies; the rule shortened the reporting window to a time-frame that many cybersecurity professionals consider too short to enable a precise understanding of exactly what data has been compromised, by whom, and in what ways.[22] Compliance with this requirement thus may complicate, rather than simplify and enable effective risk communications.

Overall, the legal landscape that shapes risk communication for digital platforms is dynamic and complicated, varies by national and sub-national geography, and is sometimes less than fully sensitive to the constraints and affordances of the underlying technologies. This can have the perverse effect of reinforcing a compliance-oriented or "box-checking" culture, since the effort and resources required to comply and reduce legal risk are so significant.

Sub-Legal Practices

Not every data security and privacy breach legally requires external notification. The regulations in place include thresholds for when notification becomes mandatory. For example, the GDPR only concerns personal data, so notification is not required under the same conditions for a breach that involves a company's intellectual property. Individuals (as compared to a supervisory authority) must be notified only if there has been exposure or unauthorized access to personal data *and* if the breach is likely to result in a high risk to those individuals' rights and freedoms.

22 Jeff Stone, "Your Network Has Been Hacked. You Have 72 Hours to Report It," *Wall Street Journal*, September 18, 2018, https://www.wsj.com/articles/your-network-has-been-hacked-you-have-72-hours-to-report-it-1537322400.

Determining exactly what meets this threshold of risk can be complicated. The sensitivity of the data, the number of people impacted, the vulnerability of those impacted, and the scope of the consequences all play a role. But if, for example, only a small number of email addresses are revealed to unauthorized people without additional sensitive data, notification to the affected individuals is likely not required.[23] In an explicit attempt to protect users from "notification fatigue," the GDPR sets a higher threshold for communicating with individuals than for notifying supervisory authorities.

Beyond what is required by the law, companies often encounter events that fall below the legal threshold and need to make (voluntary) choices about whether and how to notify users. Imagine that a third-party vendor employed by a digital platform discovers that the platform's user data has been mistakenly left unsecured for six months, and while the vendor has no evidence of malicious activity, they cannot guarantee that the information was not accessed by anyone outside their organization. It is this kind of gray area — below the legal threshold — where companies have greater autonomy. This freedom of action creates an opportunity to gain or lose user trust and reputation.

Indeed, prior to the implementation of the GDPR and other regulations, many companies already chose to disclose a broad range of privacy- and security-related breaches to users. For global companies navigating complex legal landscapes across multiple jurisdictions simultaneously, it can be especially helpful to have uniform procedures in place to facilitate communication practices.

Privacy policies have some of the same characteristics. The Pew Research Center has found that about eight in ten Americans say they are asked to agree to a privacy policy at least monthly, including one-quarter who say this happens nearly daily.[24] As is well known, privacy policies are typically lengthy documents primarily designed by and for lawyers, and are often hidden at the bottom of webpages or otherwise designed to be easily ignored. Most people have grown accustomed to agreeing to the terms of privacy policies with a click and without so much as skimming their contents. By most accounts, privacy policies are not fulfilling the role of meaningfully communicating privacy risks to users. Sixty-three percent of Americans report understanding little to nothing about the laws and regulations in place to protect their data privacy, and only 22% claim to read privacy policies in their entirety before agreeing to the

23 Debbie McElhill, "Personal Data Breaches — To Notify or Not to Notify?" Data Protection Network, October 2019, https://dpnetwork.org.uk/opinion/personal-data-breaches-to-notify-not-to-notify/.

24 Brooke Auxier et al., "Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information," Pew Research Center, November 15, 2019, https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

conditions.[25] It is hardly surprising that roughly eight in ten Americans say they have very little or no control over the data that companies collect about them.

The lack of accessibility of privacy policies is a well-known problem, and digital technology companies, among others, have worked on alternative solutions for many years. An example that is not yet well developed but has gained traction recently is the idea of "privacy nutrition labels." This concept is discussed briefly below, along with three other compelling and diverse examples of risk communication currently being used or considered by digital platforms.

Examples of Digital Risk Communication

Privacy Nutrition Labels

Just like a nutrition label on a food product, privacy nutrition labels are designed to give consumers a quick, scannable glimpse into the potential privacy risks of a digital service prior to using it, providing a proactive and standardized form of risk communication.[26] Carnegie Mellon's CyLab developed an early prototype of the privacy nutrition label in 2009 and tested its effectiveness through an online study of nearly 800 participants. The study found that "providing standardized privacy policy presentations can have significant positive effects on accuracy and speed of information finding and reader enjoyment with privacy policies."[27] By standardizing the labels, consumers can gain a more accurate understanding of how technology companies are using their data, without needing to read lengthy privacy policies.[28]

The most recent prototype of the privacy nutrition label (*Figure III*) includes both a primary and secondary layer. The primary label layer would, for example, be printed onto an IoT device, and the secondary layer would be accessible via a QR code that can be scanned by a mobile device. The secondary layer serves two purposes: curious con-

25 *Ibid.*

26 TechRadar Pro, "The Case for a Privacy Nutrition Label," *TechRadar*, January 20, 2020, https://www.techradar.com/news/the-case-for-a-privacy-nutrition-label.

27 Patrick Gage Kelley et al., *Standardizing privacy notices: an online study of the nutrition label approach*, 2010, https://dl.acm.org/doi/10.1145/1753326.1753561.

28 TechRadar Pro, "The Case for a Privacy Nutrition Label," *TechRadar*, January 20, 2020, https://www.techradar.com/news/the-case-for-a-privacy-nutrition-label.

sumers have an avenue to become better informed, and the information on the label is machine readable and can be processed by different entities.[29]

The idea of privacy nutrition labels has gained support from diverse stakeholders, including most notably the Federal Trade Commission.[30] The labels have also been subject to numerous criticisms. First, although the concise format has benefits, it can also hide meaningful nuance and context.[31] Second, it has not yet been demonstrated that privacy labels will cause people to change their behavior in any significant way.[32] Finally, there is no clear authority responsible for creating the standardized format and managing it over time. If technology companies themselves create their own labels, would they have the right incentives to provide transparent information? The installation of a neutral third-party to oversee the labels would be ideal, but could also be challenging for firms to agree upon and manage.[33]

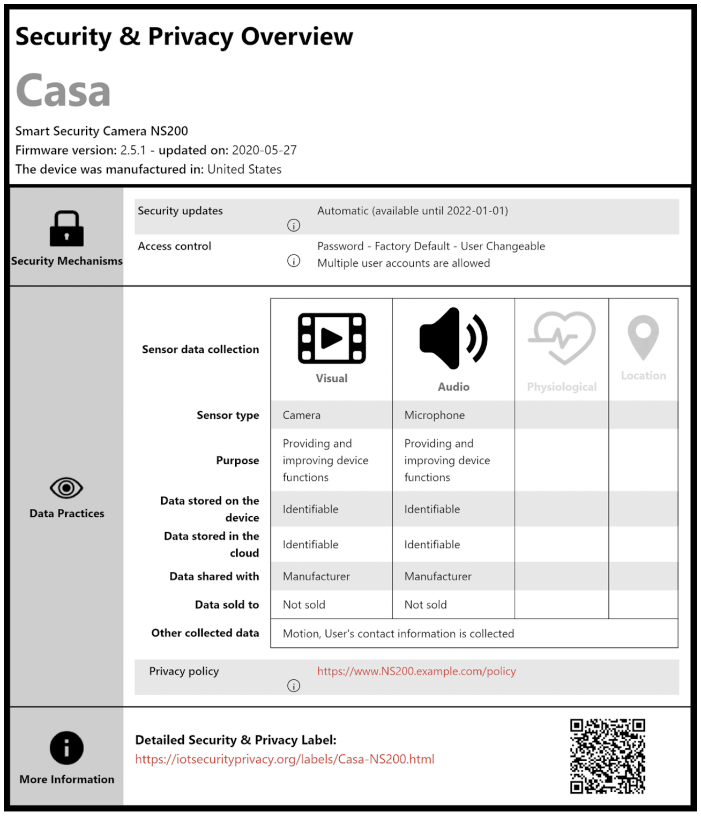


Figure III. The IoT Security & Privacy Label prototype includes information on Security Mechanisms (e.g., does the IoT device get automatic security updates, and who can access the device?), Data Practices (e.g., what data is collected, and is it shared with other companies?), and More Information (for experts or users who want more details).[34]

29 "Lily Hay Newman, "IoT Security Is a Mess. Privacy 'Nutrition' Labels Could Help," *Wired*, June 9, 2020, www.wired.com, https://www.wired.com/story/iot-security-privacy-labels/.

30 "FTC Working on Privacy "Nutrition Label"; Industry Focusing on Icons," *Inside Privacy*, October 25, 2012, https://www.insideprivacy.com/privacy-policies/ftc-working-on-privacy-nutrition-label-industry-focusing-on-icons/.

31 Helen Nissenbaum, "A Contextual Approach to Privacy Online," *Daedalus*, Fall 2011, https://www.amacad.org/publication/contextual-approach-privacy-online.

32 Kashmir Hill. "Is It Time For Privacy Nutrition Labels?" *Forbes*, https://www.forbes.com/sites/kashmirhill/2011/03/23/is-it-time-for-privacy-nutrition-labels/.

33 Pardis Emami-Naeini et al. "Ask the Experts: What Should Be on an IoT Privacy and Security Label?" *ArXiv*, February 2020, http://arxiv.org/abs/2002.04631.

34 See https://iotsecurityprivacy.org.

There are other interesting examples of how digital platforms have experimented with risk communications. For example, one approach is to integrate messages more directly into the user experience, as in the example of the Privacy Checkup Tool, which is intended to provide more proactive messaging about privacy policies.

The Privacy Checkup Tool

In January 2020, Facebook overhauled Privacy Checkup, a tool originally launched in 2014 to send a proactive message to users via their News Feeds, encouraging them to revisit their privacy choices on the social media platform.[35] The tool directs users to review four key privacy settings: what data they share with third-party applications connected to Facebook, who sees their posts, how people can find them on the platform, and how to keep their accounts secure from some simple cyberattacks. A fifth setting was later added that allows users to control the use of some of the information used to determine advertisement preferences.

Over the years, Facebook has updated the design of this tool, along with the kinds of information people can access. For example, a 2020 feature shows what information about a user has been provided to Facebook by other companies.[36] People are then given the option to delete this "Off-Facebook Activity" data from Facebook.

Other digital platforms have similarly aimed to simplify users' ability to see and change privacy and security settings. For example, since 2015, Google has offered its own Privacy Checkup tool, which encourages users to review and choose settings that control what data Google saves about users' account activity.

Another method of risk communication that has emerged from digital platforms falls on the opposite side of the spectrum. Rather than opting for a short and easily legible format, this approach is based on the publication of reports, often accompanied by websites or landing pages and extensive dissemination strategies, to share significantly more information with diverse audiences. These reports vary in the content they convey, but they are all examples of how firms voluntarily release potentially sensitive information, including how frequently government law enforcement and intelligence agencies are granted access to user data. In each

35 Paddy Underwood, "Privacy Checkup Is Now Rolling Out," Facebook Newsroom, September 4, 2014, https://about.fb.com/news/2014/09/privacy-checkup-is-now-rolling-out/.

36 Mark Zuckerberg, "Starting the Decade by Giving You More Control Over Your Privacy," Facebook Newsroom, January 28, 2020, https://about.fb.com/news/2020/01/data-privacy-day-2020.

case, the company is going beyond legal requirements in part to showcase a willingness to be transparent and to help users understand some of the complex risks they face by using the company's services.

Transparency Reports

In 2010, Google released its first "Transparency Report," which publicly disclosed data about government demands to remove content, block services, and access information about users. In the years since, transparency reports have become a relatively common practice among digital platforms, with Twitter, Microsoft, Apple, and Facebook all providing such reports by 2012. In fact, 70 companies around the world have released transparency reports to date, though the growth rate has been decreasing since 2013.[37] Transparency reports are considered to be an effective way to communicate information about user data requests by governments to digital platforms.

The kind of information included in transparency reports has shifted over the years, but the reports typically cover how companies receive, track, and respond to government requests, including where the request came from, what it included, what response it received, how the subject of the request was informed, and how many other users were impacted. Transparency reports have generally been used to build awareness and trust with users, signal company values and practices, and educate policymakers, among other uses.[38]

In some cases, transparency reports that cover government requests for user data are accompanied by what is called a "warrant canary," a statement published by the company that states they have *not* received classified government requests for data. Once the company has received such a request, they may be prohibited from disclosing information about it based on regulations such as the US National Security Letter provision.[39] By observing the removal of the warrant canary, users can infer the company is handling such a request without the company needing to explicitly say so.

37 Access Now, "Transparency Reporting Index," https://www.accessnow.org/transparency-reporting-index/.

38 "The Transparency Reporting Toolkit," Open Technology Institute at New America and the Berkman Center for Internet & Society at Harvard University, December 2016, https://cyber.harvard.edu/sites/cyber.harvard.edu/files/Transparency_Reporting_Guide_and_Template-Final.pdf.

39 Andrew E. Nieland, "National Security Letters and the Amended Patriot Act," 92 *Cornell Law Review,* 1201 (2007), https://www.lawschool.cornell.edu/research/cornell-law-review/upload/Nieland.pdf.

Safety Reports

In December 2019, ridesharing firm Uber voluntarily published its *U.S. Safety Report*, which was billed (arguably with some exaggeration) as "the first comprehensive publication of its kind to be issued by a company."[40] The 84-page report provided a unique view into the scale and scope of the most serious safety incidents that had occurred through use of the Uber platform, including traffic fatalities, fatal physical assaults, and sexual assaults. With the assertion that "people have the right to know about the safety records of the companies they rely on every day," the report was put forward as part of a proactive effort to improve transparency standards at Uber and across the industry.

The report examines data from 2017–2018, during which time an average of 3.1 million Uber trips took place every day in the United States. Preparing the report reportedly required an "intensive, nearly two-year effort." One of the challenges the company faced is that there was no uniform industry standard for counting or categorizing sexual assaults. Uber therefore partnered with the National Sexual Violence Resource Center and the Urban Institute to create an open-source classification system that could also be used by other companies.

Uber knew it was taking a gamble by publicly disclosing data about horrific incidents that had occurred while customers and drivers were using its service, and indeed the findings were disturbing to many. Even though incidents are rare overall given the large total number of rides — rarer than on many forms of public transportation, by many measures — the raw findings are nonetheless emotionally shocking and expose risks that disproportionally impact women and rarely receive widespread attention. For example, the report includes data about more than 200 rapes and around 3,000 sexual assaults each year, with an average of about eight sexual assaults every day. The response to the report illustrates one of the core challenges of risk communication: do users have the knowledge and tools they need to make sense of raw numbers like these? Should firms be responsible for providing those tools? Will the media fail both users and the firm by defaulting to sensationalism and not carefully explaining the denominator as well as the numerator in the risk equation?

40 Uber, "Uber's US Safety Report," December 2019, https://www.uber.com/us/en/about/reports/us-safety-report/.

Following the release of the report, Uber suffered short-term losses in market share value of as much as $1.4 billion.[41] Still, the findings facilitated changes throughout the industry. For its part, Uber expanded its safety team to more than 300 people, began conducting more rigorous background checks, and added an In-App Emergency Button, among other features. What we do not know is whether users were able to use the information in the report to make decisions or take meaningful action to manage the low-probability, yet high-consequence risks associated with using the Uber platform.

GAPS, COMPLEXITIES, AND TENSIONS

Transparency

Transparency is one of the most important and commonly noted features of responsible risk communication, but is not always straightforward in practice. Simply providing people with a deluge of non-contextualized raw information is rarely helpful. The "transparency paradox" refers to the dilemma that providing more information can decrease the likelihood that people read and understand the relevant information. Also well known is the negative externality of "notification fatigue," where people eventually stop noticing or acting upon notifications because they have received so many and feel powerless to manage the choices or risks they are told about.[42] Technology companies have extensive experience in trying to deal with this problem, including by trying to design notices so they are not mindlessly clicked through – for example, through design choices such as placement of notices and size and color of fonts.[43] Typically, standardizing a risk communication system across an industry serves the public's needs for transparency better than do disparate systems. However, no such standardized system currently exists in the technology industry.

41 Rex Crum, "Uber loses $1.4 billion in value after reporting thousands of sexual assaults in its rides," *The Mercury News*, December 6, 2019, https://www.mercurynews.com/2019/12/06/uber-loses-1-5-billion-in-value-after-reporting-thousands-of-sexual-assaults-in-its-rides/.

42 Ann Cavoukian, "A Discussion Paper on Privacy Externalities, Security Breach Notification and the Role of Independent Oversight," Workshop on the Economics of Information Security, University College, London, June 24, 2009, https://www.ipc.on.ca/wpcontent/uploads/Resources/privacy_externalities.pdf.

43 Erin Egan, "Communicating About Privacy: Toward People-Centered and Accountable Design," Facebook, July 2020, https://about.fb.com/wp-content/uploads/2020/07/Privacy-Transparency-White-Paper.pdf; Devdatta Akawe and Adrienne Porter Felt, "Alice in warningland: A large-scale field study of browser security warning effectiveness," {USENIX} Security Symposium ({USENIX} Security 13), 2013.

Distrust

Risk communication theorists cite the rapid development of technology, spread of social media, and proliferation of internet use as factors that have contributed to a culture of distrust.[44] Some theorists argue that this is part of a broader lack of trust between institutions and citizens, and even marks a "post-trust society."[45] Certainly, trust in technology is wavering, and is especially low for social media companies.[46] Trust is significantly influenced by perceptions about risks to users, such as threats to data privacy and security.[47] Failing to communicate about these threats can be seen as obfuscation and tends to increase the trust deficit.[48] Re-establishing trust is notoriously difficult and requires active engagement with users and other stakeholders.[49]

Scalability

Digital technologies are capable of scaling to millions (or billions) of users around the world. Effective communication at that scale is significantly more complex, as it requires awareness of different languages, legal frameworks, and cultural norms. People interact with technologies in different ways, and face disparate impacts from threats. While legal teams are used to accounting for regional differences, the role of psychology, sociology, and anthropology in understanding the needs of diverse users has generally been under-appreciated in practice. No risk communication can be unbiased, but developing greater understanding of the psychological, social, and environmental influences and constraints of an audience improves a company's ability to effectively communicate. There is also an ongoing tension between designing standardized communications that support interpretability and evaluation while also enabling adaptability and specificity that can more appropriately reach diverse audiences.

44 Paul Slovic. "Perceived Risk, Trust, and Democracy." *Risk Analysis*, vol. 13, no. 6, Dec. 1993, pp. 675–82. DOI.org (*Crossref*), doi:10.1111/j.1539-6924.1993.tb01329.x; Dominique Brossard, "New Media Landscapes and the Science Information Consumer." *Proceedings of the National Academy of Sciences*, vol. 110, no. Supplement_3, Aug. 2013, pp. 14096–101. DOI.org (*Crossref*), doi:10.1073/pnas.1212744110.

45 Ragnar Löfstedt. *Risk Management in Post-Trust Societies.* Earthscan, 2009.

46 Sanjay Nair, "Trust in Tech is Wavering and Companies Must Act," Edelman Research, April 8, 2019, https://www.edelman.com/research/2019-trust-tech-wavering-companies-must-act.

47 Carlos Flavián and Miguel Guinalíu, "Perceived security and privacy policy: Three basic elements of loyalty to a web site," *Industrial Management & Data Systems*, June 1, 2006, https://www.emerald.com/insight/content/doi/10.1108/02635570610666403/full/html.

48 Ann Cleaveland, Jessica Newman, and Steven Weber, "The Art of Communicating Risk," *Harvard Business Review*, September 24, 2020, https://hbr.org/2020/09/the-art-of-communicating-risk.

49 Tech executives commonly quote Warren Buffet's saying that "it takes years to build a reputation and seconds to destroy it."

In practice, this tension has implications for how risk communications should be structured in global firms, which may, for example, choose to develop localized risk communications expertise for the countries where they operate. Different firms will solve for this in different ways, but ideally, risk communications should be pushed down through organizations into product development, and relevant social science skills should be integrated into regional and global product teams.

Uncertainty

Uncertainty — in this instance, the inability to assign a probability distribution to outcomes — is a core feature of the overall risk landscape that affects users on digital platforms. The history of financial markets is in large part a story of translating uncertainty into quantified measures of risk that can be hedged and traded, as well as explained more readily. But privacy and security risks almost always involve uncertainty because of how data can be re-used for different purposes, and this kind of uncertainty is difficult to communicate and understand, much less hedge and trade.[50] Nor do providers of digital products yet have a standardized framework for communicating uncertainty with users and customers; many firms are hesitant to disclose uncertain information or will disclose uncertainty in an ambiguous way, leading to confusion or misinterpretation.[51] Upstream public engagement, in which the uncertain futures of a technology product or service are discussed with members of the community in an ongoing dialogue, is one approach that shows some promise in increasing the public's risk fluency, and demonstrating what meaningful and transparent communication looks like under uncertainty.[52] But the difficulty of translating uncertainty into quantifiable measures of security and privacy risk is a tension that communicators will likely be faced with for a long time to come.

50 Ann Cleaveland, Jessica Newman, and Steven Weber, "The Art of Communicating Risk," *Harvard Business Review*, September 24, 2020, https://hbr.org/2020/09/the-art-of-communicating-risk.

51 Baruch Fischhoff. "Communicating Uncertainty Fulfilling the Duty to Inform." *Issues in Science and Technology*, vol. 28, no. 4, 2012, pp. 63–70, JSTOR.

52 Nick Pidgeon, Tee Rogers-Hayden, "Opening up nanotechnology dialogue with the publics: Risk communication or 'upstream engagement'?" *Health, Risk & Society*, 13698575, June 2007, Vol. 9, Issue 2.

Part II: Theory and Cross-Industry Learning

THE SCIENCE OF RISK COMMUNICATION

The science of risk communication is situated at the intersection of decision science, psychology, sociology, and communications. People have been communicating about hazards, analyzing risk, and labeling harmful materials for centuries. The discipline focuses on studying ways in which organizations and individuals can implement more effective risk communication strategies while engaging with key stakeholders and the people impacted by those strategies. This includes creating shared understanding of the facts and helping people make sound choices by informing them about benefits, risks, and costs of their decisions, as well as arming them with tools and techniques to make use of that information effectively.

Fischhoff (of Carnegie Mellon University) describes effective risk communication as including the following four steps:[53]

1. Identify the science most relevant to the decisions people face
2. Determine what people already know
3. Design communication to fill the critical gaps
4. Evaluate adequacy and repeat as necessary

This is a simple discipline in its basic structure that is rarely implemented on a consistent basis, in part because it is much harder to fulfill in practice and in part because it often requires re-framing relationships between speakers, listeners, and decision-makers about risk. Fischhoff also notes that the theory of risk communication has evolved over time, and arguably has moved through eight developmental stages over the past twenty years.[54] These stages have

53 Baruch Fischhoff, "The Sciences of Science Communication." *Proceedings of the National Academy of Sciences*, vol. 110, no. Supplement_3, Aug. 2013, pp. 14033–39. DOI.org (*Crossref*), doi:10.1073/pnas.1213273110.

54 Baruch Fischhoff, "Risk Perception and Communication Unplugged: Twenty Years of Process," *Risk Analysis*, vol. 15, no. 2, 1995, pp. 137–45. *Wiley Online Library*, doi:10.1111/j.1539-6924.1995.tb00308.x.

advanced the field from a relatively simple focus on communicating numbers to decision-makers (e.g., number of fatalities that a given technology produces in an average year) toward a more complex notion of relationship building. This is informed by the realization that it is not possible to inform decision-making without a textured understanding of the people who are making those decisions. This led to a key theoretical shift away from top-down authoritative risk communication strategies to an approach that is more transparent and collaborative, with increasing recognition of the value of citizen engagement in scientific knowledge production (not simply reception and processing). Thus, for many organizations, moving toward effective risk communication requires a cultural shift as well.

Relevant Themes in Risk Communication Theory

A major focus of risk communication literature centers on the relationships between decision-makers, "expert authorities," and other relevant stakeholders and publics. In order to know what matters to anyone who receives a given risk communication, specialists suggest using mental models and influence diagrams.[55] Mental modeling involves understanding the bases of peoples' reactions to different messaging, and requires significant interviewing and surveying to be done successfully. Influence diagrams are another way of understanding the roles and involvement of individuals and groups in a given risk situation. Both of these techniques are reiterated throughout risk communication theory as suggestions to be implemented across firms and institutions. It would be convenient if there were inductively-derived, simple, standard models, but the factors that go into people's rationale and decision-making processes are numerous and surprisingly difficult to generalize across time, communities, and cultures.[56] People tend to perceive risks in highly subjective and context-dependent ways, and through various judgment heuristics that can skew perceptions dramatically.[57]

Another relevant theme of risk communication theory is the importance of building and maintaining trust. Numerous studies have examined how this can be done, but a standard practice for facilitating trust is two-way communication.[58] This involves careful and sustained listening among a wide range of potential audiences. Uncertainty also plays an important role, because risks and potential risks often include varying levels of uncertainty that are not always

55 M. Granger Morgan, et al. *Risk Communication: A Mental Models Approach*. Cambridge University Press, 2002.

56 Paul Slovic. *The Perception of Risk*. Routledge, 2016.

57 Daniel Kahneman, et al., editors. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, 1982.

58 Baruch Fischhoff, "When Assessing Novel Risks, Facts Are Not Enough." *Scientific American*, doi:10.1038/scientificamerican0919-74. Accessed 9 July 2020; National Research Council (U.S.), editor. *Improving Risk Communication*. National Academy Press, 1989.

clearly communicated among multiple stakeholders. A 2010 study found that experts who expressed uncertainty were seen as more credible to a lay audience than those who did not.[59] Interestingly, sources who were seen *a priori* as having lower expertise were perceived as *more* credible if they expressed certainty.

Critiques and Evolution of Thought

One critique of the field of risk communication is that it can be seen as a form of manipulation or social engineering. While it is the goal of the communicator to provide people with the information they need to make an informed decision or take an action, no communicator can be unbiased in doing that. All risk communications are constructed, and relevant choices are made about how to present the dangers and available options. These choices add up to paternalistic guidance of a kind, whether intentional or not. It is impossible to construct risk communication that serves "the public" alone, and since risk communicators "do not fund research to study themselves, they remain invisible in most risk research."[60]

Science and Technology Studies (STS) scholars have also advocated for moving beyond the notion of two-way communication toward a more comprehensive model of upstream public engagement, a reasonable definition of which is "dialogue and deliberation amongst affected parties about a potentially controversial technological issue at an early stage of the research and development process and in advance of significant applications or social controversy."[61] STS theorists typically reject the traditional notion that a layperson suffers simply from a knowledge or understanding deficit relative to experts, in favor of a more social explanation that is about fostering democratic governance of science, as well as a "reflexive" learning process in science institutions and policies. But while these critical studies point to important normative considerations for practice, they do not generally provide practical guidance that can be immediately applied to improve the practices of risk communicators right now.

59 Uma R. Karmarkar, Zakary Tormala, "Believe Me, I Have No Idea What I'm Talking about: The Effects of Source Certainty on Consumer Involvement and Persuasion." *Stanford Graduate School of Business*, https://www.gsb.stanford.edu/faculty-research/publications/believe-me-i-have-no-idea-what-im-talking-about-effects-source. Accessed 9 July 2020.

60 Roger E. Kasperson and Pieter Jan M. Stallen. *Communicating Risks to the Public: International Perspectives*, Springer, 1991.

61 Nick Pidgeon and Tee Rogers-Hayden, "Opening up nanotechnology dialogue with the publics: Risk communication or 'upstream engagement'?" *Health, Risk & Society*, 13698575, Jun2007, Vol. 9, Issue 2.

LESSONS FROM OTHER SECTORS

Experts grapple with the challenges and complexities of risk communication across a wide variety of fields. Allocating the appropriate resources, designing protocols, and understanding the public's perception of a problem are all pervasive issues that characterize communicating risks across industries. From natural disasters to prescription medication, industries struggle with communicating risk, and many consult with the risk communication theory discussed previously. While there are gaps between theory and practice, the following mini-case studies demonstrate some worthwhile takeaways from risk communication practice in the field. These are all relevant examples that can be applied to digital risk communication, and provide inspiration for emerging practices.

Natural Disaster Response and Recovery: 2011 Mississippi River Flood

In 2011, following a winter of heavy snowfall, the US Army Corps of Engineers realized that a spillway along the Mississippi River would need to be opened, flooding rural Louisiana farmland in order to prevent the flooding of major urban centers, and requiring the evacuation of rural residents. Despite having not flooded these residents' properties since the 1970s, the Army Corps had sent residents a letter every year reminding them of the potential risk. When the time came for the Army Corp to facilitate an evacuation, residents were fully informed and aware of the situation. The preemptive and regularly scheduled communication of risk allowed for a smoother transition when the Army Corp had to prepare residents for an actual flood.

The letters sent by the Army Corps were also quoted and reiterated by the media before the time of the flood. This dual approach reinforced the messages sent by the Army Corps in their letters: "Louisiana residents are being warned today: The Army Corps of Engineers will open the Morganza spillway along the Mississippi River by Sunday, flooding millions of acres of rural farmland and sparing big cities like Baton Rouge and New Orleans."[62] By designing letters to be read by multiple audiences (residents and the media alike), the Army Corps was able to effectively distribute and reinforce their message in a timely manner.

62 Jim Avila, Yunji de Nies, and Enjoli Francis, "Mississippi Floods: Spillway to Be Opened in Louisiana," *ABC News*, May 12, 2011, https://abcnews.go.com/US/mississippi-river-floods-spillway-opened-louisiana/story?id=13600771. Accessed 17 July 2020.

Despite the enormous economic damage of the 2011 Mississippi Flood, the Army Corps' approach to communicating risk to the public has been deemed effective by many experts.[63] By communicating continuously for years, being strategic about the purpose, structure, and organization of letters, and writing the letters with the intention that they would be read by multiple readers, the Army Corp successfully relayed relevant information prior to the opening of the floodway.[64,65] This example highlights the importance of early and continuous communication (not just at the moment when a crisis occurs), the thought and strategy necessary to compose a successful communication, and the importance of writing to multiple audiences and utilizing the media as a tool to help spread the message.

Medical Products: Drug Facts Box

American consumers are exposed to billions of dollars worth of prescription drug advertisements each year, messages that typically lack key information about the effectiveness of a drug.[66] Marketers of course focus on generating demand for pharmaceuticals, rather than informing the public about potential risks.[67] Today, pharmaceutical companies draft their own product labels and the FDA approves them. Although this process is regulated, the structure introduces a conflict of interest, because the entity that first decides what is deemed as essential consumer knowledge is also marketing the drug. In response, experts have proposed an additional standardized method of communicating to consumers: the drug facts box. This box consists of "a one-page summary of benefit and harm data for each indication of a drug." National studies have demonstrated that the drug facts box improves consumer decision-making and leads to a more informed public.[68]

63 Carolyn Boiarsky, *Risk Communication and Miscommunication: Case Studies in Science, Technology and Community Organizations,* Chapter 2: "Effective Discourse Strategies," University Press of Colorado, 2016.

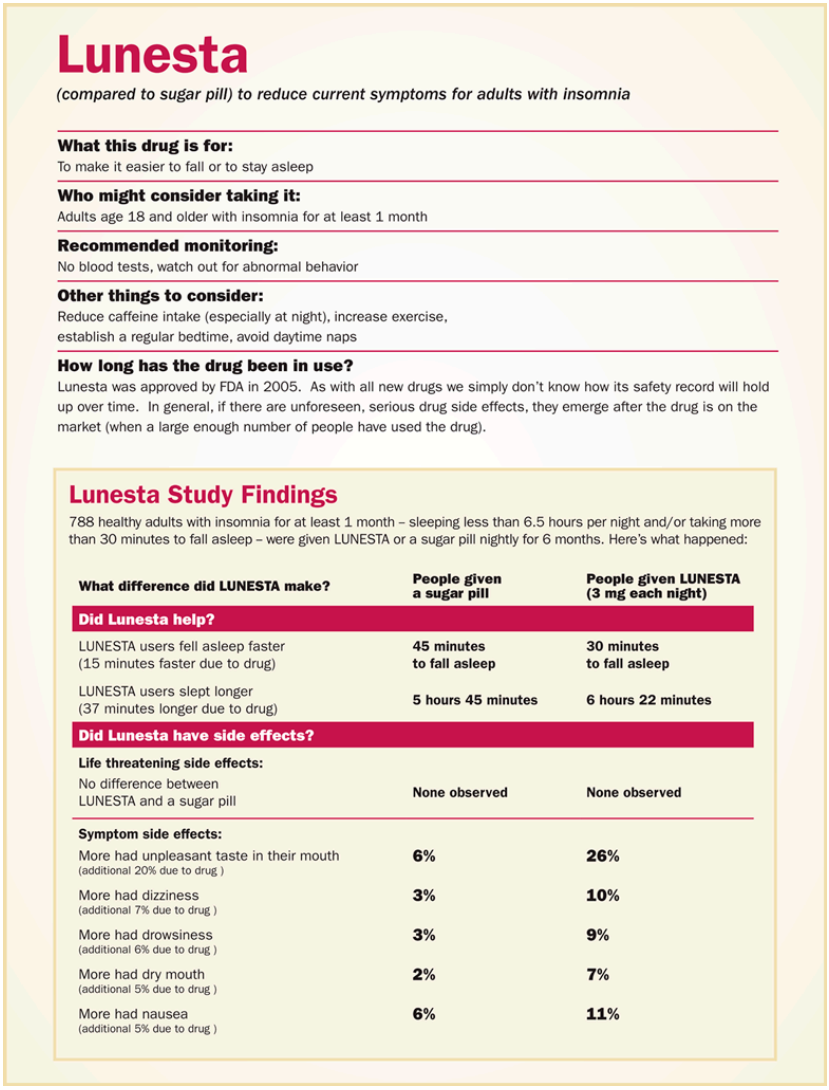
64 Boiarsky notes that "shortly after the [final] letter was mailed, local officials held a series of meetings for local residents. . . . By May 14, when the floodway opened, all but the most stubborn of residents had departed the region safely [and] although they had not wanted to leave, few had complaints" (32).

65 Other examples from natural disaster warnings have brought to light the potential of "false alerts" to negatively impact the public's trust in the system, hinging on whether the receiver believed the alert was justified. See J.E. Trainor, et al. "Tornadoes, social science, and the false alarm effect," *Weather, Climate, and Society*, 7 (4) (2015), pp. 333–352.

66 Lisa M. Schwartz and Steven Woloshin, "Communicating Uncertainties about Prescription Drugs to the Public: A National Randomized Trial," *Archives of Internal Medicine*, vol. 171, no. 16, Sept. 2011, pp. 1463–68. *PubMed*, doi:10.1001/archinternmed.2011.396.

67 M.F. Hollon, "Direct-to-Consumer Marketing of Prescription Drugs: Creating Consumer Demand," *JAMA*, vol. 281, no. 4, Jan. 1999, pp. 382–84. *PubMed*, doi:10.1001/jama.281.4.382.

68 Lisa M. Schwartz, et al, "Using a Drug Facts Box to Communicate Drug Benefits and Harms: Two Randomized Trials," *Annals of Internal Medicine*, vol. 150, no. 8, Apr. 2009, pp. 516–27. *PubMed*, doi:10.7326/0003-4819-150-8-200904210-00106.



Lunesta

(compared to sugar pill) to reduce current symptoms for adults with insomnia

What this drug is for:
To make it easier to fall or to stay asleep

Who might consider taking it:
Adults age 18 and older with insomnia for at least 1 month

Recommended monitoring:
No blood tests, watch out for abnormal behavior

Other things to consider:
Reduce caffeine intake (especially at night), increase exercise, establish a regular bedtime, avoid daytime naps

How long has the drug been in use?
Lunesta was approved by FDA in 2005. As with all new drugs we simply don't know how its safety record will hold up over time. In general, if there are unforeseen, serious drug side effects, they emerge after the drug is on the market (when a large enough number of people have used the drug).

Lunesta Study Findings

788 healthy adults with insomnia for at least 1 month – sleeping less than 6.5 hours per night and/or taking more than 30 minutes to fall asleep – were given LUNESTA or a sugar pill nightly for 6 months. Here's what happened:

What difference did LUNESTA make?	People given a sugar pill	People given LUNESTA (3 mg each night)
Did Lunesta help?		
LUNESTA users fell asleep faster (15 minutes faster due to drug)	45 minutes to fall asleep	30 minutes to fall asleep
LUNESTA users slept longer (37 minutes longer due to drug)	5 hours 45 minutes	6 hours 22 minutes
Did Lunesta have side effects?		
Life threatening side effects: No difference between LUNESTA and a sugar pill	None observed	None observed
Symptom side effects:		
More had unpleasant taste in their mouth (additional 20% due to drug)	6%	26%
More had dizziness (additional 7% due to drug)	3%	10%
More had drowsiness (additional 6% due to drug)	3%	9%
More had dry mouth (additional 5% due to drug)	2%	7%
More had nausea (additional 5% due to drug)	6%	11%

Figure IV: Source: "Inside the Drug Facts Box," Dartmouth Medicine, Spring 2008, Courtesy of Steven Woloshin, founder of the Lisa Schwartz Foundation.

The main feature of the drug facts box is a data table that outlines the quantified risks of various outcomes with and without the drug (see Figure IV). The intentions of the drug facts box are threefold: it assists physicians by providing a quick summary of the benefits and side effects of a drug; it educates consumers by filling in gaps that exist in drug advertisements and medication guides; and it improves physician–patient communication by facilitating evidence-based discussions about drugs. By intentionally opening up avenues for better communication between doctors and patients, the drug facts box helps foster trust and transparency, two pillars critical to effective risk communication. Due to the complexity of determining what

information to include in the drug facts box, researchers have developed a handbook to support decisions regarding what information is most important to consumers and doctors.[69]

Some of the transferable principles of the design of the drug facts box include the standardized way of highlighting risks and benefits of a product — in many cases comparing a particular drug to known alternatives — and the inclusion of empirical evidence to support the metrics provided.

Public Health: Communications about Zika Virus

The responses to Ebola, H1N1, Zika, and now COVID-19 have provided poignant and timely examples of risk communication in the field of public health. After the 2016–2017 Zika virus outbreak, experts conducted a study to examine the risk communication practices of 13 public health policymakers and practitioners, 10 public information officers, and five vector-control officials who were responsible for distributing information during the outbreak.[70] The study found that understanding the macro-environment of a community — the economic, political, and social forces that influence both trust and response to a crisis — was crucial for effectively communicating risk.

Mistrust of authorities, specifically the government, presented an enormous challenge to effective risk communication. In the case of Zika, language barriers caused issues in relaying trusted risk messages to communities. Even without a crisis, longstanding mistrust of the government can prevent effective risk messaging. A potential solution proposed in the study is to expand communication efforts to mobilize non-governmental organizations, the media, and private companies. The media and social media of course now play an outsized role in the public's perception of a crisis or outbreak. In the case of Zika, the media was helpful in spreading the word about the virus at the beginning of the outbreak, but as soon as media attention faded, the public was more likely to downplay the severity of the virus. Social media, however, served as a cost-effective way of reaching a large audience, and the social media

69 Lisa M. Schwartz and Steven Woloshin., "The Drug Facts Box: Improving the Communication of Prescription Drug Information," *Proceedings of the National Academy of Sciences*, vol. 110, no. Supplement 3, Aug. 2013, pp. 14069–74. www.pnas.org, doi:10.1073/pnas.1214646110.

70 Tara Kirk Sell, et al. "A Public Health Systems View of Risk Communication About Zika," *Public Health Reports*, April 2020. Sage CA: Los Angeles, CA, journals.sagepub.com, doi:10.1177/0033354920912215.

campaigns implemented during the Zika outbreak have been cited as effective information distribution channels.[71]

The number of sociological, psychological, environmental, and economic forces at play during a public health emergency make risk communication a particularly complex task. As in all fields, there is no one way to communicate risk effectively; however, understanding the macro-environment of a community by collaborating across agencies, and working to actively build trust, are valuable ways to approach this type of scenario. Collaboration across agencies must be implemented prior to a crisis, so risk communication should be incorporated into an organization's ongoing strategy.

Food and Beverage Industry: Front-of-Pack Labeling

Nutrition fact labels have long been considered an industry best practice for communicating the nutritional value of food. Yet these labels can be confusing or difficult to understand, especially among those with low health literacy. The FDA dedicates a page of its website to nutrition fact literacy, which extensively demonstrates the complexity of the labeling device.[72] A simpler, auxiliary measurement device is the front-of-pack label (FOP), which has emerged as a complement to nutrition fact labels to give consumers a sense of the healthfulness of a product by glancing at the front of a product.

While the effectiveness may vary from country to country, a study conducted with adults from Canada, the U.S., Australia, and the UK found that front-of-pack labeling was helpful for consumers with different levels of health literacy and had a strong influence on consumers' perceptions.[73] In the United States, the FDA has yet to adopt front-of-pack labeling as a standardized system; however, manufacturers and nonprofits globally have developed various systems, such as the Smart Choices Program, Guiding Stars, and the British traffic light system.[74] The challenge of not having a government-led standardized system is that the nutrients selected for inclusion in a program might not present a full picture of the nutrient

71 *CERC Manual | Crisis & Emergency Risk Communication (CERC)*. 25 Feb. 2020, https://emergency.cdc.gov/cerc/manual/index.asp.

72 Center for Food Safety and Applied Nutrition, "How to Understand and Use the Nutrition Facts Label," *FDA*, Mar. 2020. www.fda.gov, https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label.

73 Samantha Goodman et al., "The Impact of Front-of-Package Label Design on Consumer Understanding of Nutrient Amounts," *Nutrients*, 2018 Nov; 10(11): 1624, ncbi.nlm.nih.gov/pmc/articles/PMC6266389/.

74 Center for Food Safety and Applied Nutrition, "Background Information on Point of Purchase Labeling," *FDA*, Feb. 2020. www.fda.gov, https://www.fda.gov/food/food-labeling-nutrition/background-information-point-purchase-labeling.

profile of a product. Yet the front-of-pack labeling system serves as an easy-to-interpret identifier of how healthy a product is. While a uniform standard might be more effective, disparate systems are at least a step in the right direction when it comes to informing consumers.

Unlike the drug facts box case study, where a more nuanced table of information serves the public interest better, the front-of-pack label is small and succinct. Clearly, the context of risk communication matters, and differences between the back of a prescription drug or the back of a soft drink need to be addressed. Everyday products, such as food purchased at a grocery store, may need simplified risk labeling systems in order to appeal to a broader audience.

Transportation: Aviation Safety Reporting System

The Aviation Safety Reporting System (ASRS) was established in April 1976 in response to the 1974 crash of a United Airlines flight. Prior to the establishment of the ASRS, there was no standardized channel of communication available for sharing information about critical incidents and safety risks between airlines. The Aviation Safety Reporting Program (a precursor to the ASRS) was established to encourage members of the aviation industry to provide the Federal Aviation Administration (FAA) with timely reports about critical incidents. NASA eventually stepped in to improve the program and designed the modern ASRS system. The ASRS is an inter-organizational risk communication system for an entire industry. Aviation mechanics, air traffic controllers, pilots, crewmembers, and observers all provide input. The ASRS sends out alert bulletins and time-critical notices across the industry to warn about risks and dangers in the air and on the ground.[75] Internally, the system serves to effectively communicate dangers and hazards, and externally it serves as a confidential reporting model for other industries.[76]

The slogan on the ASRS website, maintained by NASA, is "Confidential. Voluntary. Non-Punitive."[77] These words instill trust by ensuring that information provided to the ASRS is kept confidential and that no punishment will be given to those who voluntarily submit information. An easily accessible, secure form to submit a report is available via the ASRS website. A note about security at the bottom of the web page reads, "Security features include: encryption

75 Phillip K. Tompkins, *Managing Risk and Complexity Through Open Communication and Teamwork*, Purdue University Press, 2015.

76 *Aviation Safety Reporting System Program Briefing*, https://asrs.arc.nasa.gov/overview/summary.html.

77 *ASRS - Aviation Safety Reporting System*. https://asrs.arc.nasa.gov/. Accessed 21 July 2020.

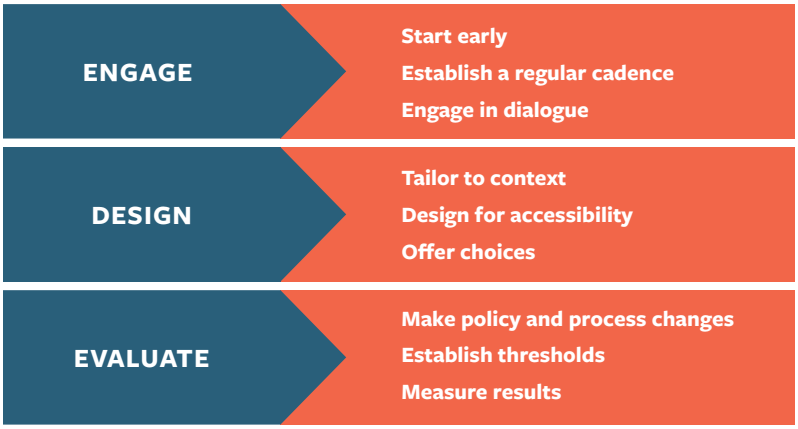
of your report form during data transmission and storage, and multiple layers of firewalls and security devices."[78] NASA's commitment to security in collecting aviation safety reports enhances the reliability of the system.

The ASRS is designed for use and consumption by professionals, not the general public, but it still demonstrates several important components of effective risk communication: feedback, risk analysis, and independence. The multi-dimensional system allows for multiple inputs to direct the content of alert bulletins. While the information is eventually synthesized by the NASA-run program, all inputs are accepted and analyzed for accuracy. Finally, the decision to have NASA run the ASRS eliminates the bias that the FAA would have imposed if the ASRP had not been disbanded. Sometimes the most effective risk communication must come from an independent third-party in order to be trusted.

78 *ASRS - Aviation Safety Reporting System - Electronic Report Submission*. https://asrs.arc.nasa.gov/report/electronic.html. Accessed 21 July 2020.

Part III: A Practical Agenda Forward

The Risk Communication Roadmap for Digital Platforms, described below, consolidates insights from these mini-cases, from expert interviews, and from a multi-stakeholder workshop held in the spring of 2020 at the University of California, Berkeley. It also incorporates lessons described throughout this paper from the theory of risk communications, best practices from other sectors, and the real-world experiences of practitioners. Firms may use this roadmap to integrate risk communication into their design and strategy, identify new opportunities for effective engagement with customers, and ultimately build more sustainable and resilient risk communication practices for the long term.



RISK COMMUNICATION ROADMAP FOR DIGITAL PLATFORMS

This roadmap simultaneously highlights actions that digital platforms can take and identifies practices that users may come to expect from responsible actors.

Following the roadmap requires that firms not silo risk communication within legal compliance or public relations departments. It requires proactive and integrated action from across a

company, including legal, policy, communications, engineering, and UI/UX design departments. This approach is necessary to ensure that users' needs and preferences are integrated into key design and policy decisions.

The roadmap helps to establish shared expectations between digital platforms and their users, and to chart a path forward based upon mutual responsibility and willingness to engage. Users should not be passive participants in this process, but should voice their needs by engaging in or advocating for proactive dialogue and opportunities for change.

1. **ENGAGE.** *Build trusting relationships with users based on transparent, comprehensive dialogue to facilitate effective risk communication.*

    a. **Start early.** Platforms should strive to establish communication channels with users before something goes wrong, and proactively communicate plans for risk prevention, preparedness, and mitigation. Robust foresight practices and risk analysis provide a basis for identifying potential threats and determining what and when to communicate. If a specific risk is reasonably expected to be on the horizon, firms should take advantage of the knowledge to warn users in advance and provide them with options about how to prepare.

    b. **Establish a regular cadence.** Communication with users should take place regularly to track shifting needs and perceptions, and to avoid information voids that will be filled by less reputable sources. People should be able to expect high-quality communication at key moments, rather than just one-off bad news. It is better to be upfront about what is known and unknown than to wait until everything is known. The appropriate cadence of communication should be regularly reassessed to provide transparency without overwhelming users. This boundary can be expected to shift depending on the circumstance, with more communication expected in times of upheaval or unease.

    c. **Engage in dialogue.** Effective risk communication demands the establishment of a variety of forums for constructive dialogue between platforms and users. Firms should not only inform users about risks, but also learn from users about what risks mean in practice, what are the most effective means of addressing them, and how communications are received. This demands an ongoing process of dialogue established through community meetings, online forums, surveys, and other mechanisms.

2. **DESIGN.** *Create accessible, informative, and actionable communication formats to enable effective risk communication.*

    a. **Tailor to context.** There is no single "public." Different stakeholders, communities, and individuals are likely to require different kinds of information and resources in order to benefit from communications and to make available options more feasible. Communication strategies have to address linguistic, cultural, and political realities, especially when operating in countries or contexts that are less well understood by a company's staff.

    b. **Design for accessibility.** Designing accessible, user-friendly risk communication means not simply inundating people with numbers, facts, and information, but also contextualizing that information so that it is relatable, based on the environment and existing understandings. Additionally, information should be displayed with a format and structure that are clear and effective, as determined through decision-science insights and user testing. Firms should use or create standardized formats where appropriate, to promote consistency across platforms and services and support informed evaluation.

    c. **Offer choices.** Risk communication efforts can start with the most important information first, but they should provide straightforward pathways to additional layers of information for people who want to know more about the details. Communications should also include options about what users can do and how to manage potential risks, from increasing security practices to being able to leave a platform if desired. These options should additionally provide resources for physical, mental, and emotional safety where relevant and feasible.

3. **EVALUATE.** Establish processes for risk communication in advance, and create metrics to assess effectiveness and ensure the resilience of risk communication efforts.

    a. **Make policy and process changes.** Risk communication should be an expected part of the business strategy and design process for digital platforms. Information flows and escalation pathways should be established well before they are needed to ensure the correct people and departments

know what actions to take and when. Messages can be pre-determined for particular situations using decision trees and scenario matrices.[79] Companies must also protect the ability of insiders and outsiders to be whistleblowers and report potential problems if necessary. Such policy and process changes, together with strong leadership, will help facilitate the changes required to shift a risk communication culture that is strictly compliance-focused toward becoming fully integrated within design and strategy.

b. **Establish thresholds for communication.** Establishing useful triggers and thresholds is a critical element of a company's risk communication process. Such thresholds should not only be based on the number of people affected, but also on relevant levels of risk and degrees of impact, and whether vulnerable communities are impacted. Uncertainty should not be used as an excuse to forgo communication.

c. **Measure results.** The effectiveness of risk communication efforts can be tested ahead of time, and should be analyzed after the fact using consistent, meaningful metrics that include not only whether the communication was "opened," but also whether people were able to take actionable steps, receive all the information they required, and communicate with the company about any issues that came up, as well as whether there were any disparate impacts identified. Ideally, metrics can be shared across the industry to create common evaluation frameworks. Lastly, it is important to engage in fine-tuning during a communication effort, and to conduct deliberate post-mortems to learn from experiences and adjust accordingly.

79 S.K. McBride, et al. "Developing post-alert messaging for ShakeAlert, the earthquake early warning system for the West Coast of the United States of America." *International Journal of Disaster Risk Reduction*, Volume 50, 2020, 101713, ISSN 2212-4209, https://doi.org/10.1016/j.ijdrr.2020.101713.

Conclusion

The dominant model of risk communication for digital platforms is still based upon a template in which compliance with narrow security and privacy breach notification laws serve as the primary motivation and yardstick of "success." Some firms have moved beyond this framework and provide additional mechanisms for communication in cases where unauthorized exposure and possible exploitation of user data falls below the legal threshold. However, there has not yet been a sustained effort across the industry to create a shared framework that users can come to rely upon and better understand over time. Part of the challenge comes from the common but overstated belief that each platform faces unique risks and communication challenges. In fact, our research has highlighted a surprising number of similarities and shared considerations. The industry has also underutilized relevant and usable insights from the science of risk communication.

This paper helps to fill these gaps by situating what is often seen as a narrow legal issue of disclosure and notification within the broader frame and scientific literature of risk communication. Current practices by digital platforms are described, including several specific examples and common challenges, including distrust, transparency, scalability, and uncertainty. The paper then assesses practices from other domains, as well as insights from multidisciplinary experts. These highlight many lessons and principles, which are consolidated into an actionable roadmap for digital platforms and their users.

Risk communication is not a panacea. To begin with, it relies upon a comprehensive assessment of risks and associated uncertainties, which pose significant challenges of their own. And communication by itself cannot guarantee that there are meaningful ways to reduce the risk or mitigate its effects. Risk communication can best be understood as one component of a larger constellation of practices, policies, and legislation that work together to hold digital platforms accountable and incentivize responsible practices that improve the risk-benefit ratio that users and consumers face.

How a company chooses to handle risks to its users says a lot about its priorities. When unanticipated events are handled well, it can serve to increase user trust. Implementing a thoughtful strategy for risk communications provides a valuable opportunity for platforms, as well as for users who can become more informed and empowered to take actions. Thoughtful communications during difficult times provide a way to credibly showcase a commitment to

user safety and wellbeing. In an environment of expanding digital risks and growing distrust of technology companies, a shared roadmap for risk communications can leverage the interdependence of many digital platforms, help "raise all boats," and support an improved environment for both users and digital platform firms.

About the Authors

ANN CLEAVELAND is the Executive Director of the Center for Long-Term Cybersecurity. She was previously the Senior Director of Strategic Planning at the ClimateWorks Foundation, where she led multiple initiatives focused on supporting a large, philanthropic collaborative in a more strategic, effective, and science-based response to global climate change.

STEVEN WEBER is Faculty Director of the Center for Long-Term Cybersecurity and Professor at the University of California, Berkeley School of Information. His research, teaching, and advisory work focus on the political economy of knowledge-intensive industries, with special attention to health care, information technology, software, and global political economy issues relating to competitiveness.

JESSICA NEWMAN is a Research Fellow at the UC Berkeley Center for Long-Term Cybersecurity, where she conducts research on digital governance and the security implications of artificial intelligence, and is the Program Lead for the AI Security Initiative. She is also an AI Policy Specialist with the Future of Life Institute and a Research Advisor with The Future Society.

GRACE GORDON is Master of Development Practice student at UC Berkeley focusing on tech policy and cybersecurity. She was a 2020 Summer Graduate Student Researcher at the Center for Long-Term Cybersecurity.

Acknowledgments

The Center for Long-Term Cybersecurity (CLTC) would like to thank participants in the February, 2020 workshop "Risk-based communication for digital privacy and security" for contributing their ideas and expertise to this report. Special thanks also to Baruch Fischhoff, Anne Wein, Chris Hoofnagle, Kristin Berdan, and Chuck Kapelke for their advice and feedback.

This project is made possible by a grant from the William and Flora Hewlett Foundation with additional funding from Facebook in support of independent academic research.

CLTC

Center for Long-Term Cybersecurity

UC Berkeley

Center for Long-Term Cybersecurity
cltc.berkeley.edu
@CLTCBerkeley