9

Risk and the Pluralism of Digital Human Rights Fact-Finding and Advocacy

Ella McPherson

I INTRODUCTION[1]

The rise of information and communication technologies (ICTs) has captivated many human rights practitioners and scholars. Particular interest, mine included, is focused on the potential of using ICTs to support the pluralism of human rights fact-finding and advocacy.[2] In theory, now anyone with a cell phone and Internet access can document and disseminate evidence of human rights abuses. But what happens when this theory is put into practice?[3] What happens when ICTs are adopted in empirical realities shaped by unique contexts, distributions of resources, and power relations?[4] I will argue that, while the rise of ICTs has certainly created new opportunities, it has also created new risk – or negative outcomes – for human rights practitioners. This risk is silencing, and unequally so.

In this chapter, I focus on human rights fact-finding and advocacy from the perspective of practitioners at human rights NGOs, while acknowledging that the range of practices and actors involved in human rights work is much broader.[5] These practices form a communication chain: information moves from witnesses on the ground to human rights practitioners during fact-finding, who gather and evaluate this information for evidence of violations. This evidence is then packaged and communicated to audiences such as journalists, policy-makers, and publics as

[1] This work was supported by the Economic and Social Research Council (grant no. ES/K009850/1) and the Isaac Newton Trust.

[2] I lead a project at the University of Cambridge called "The Whistle," which is a digital app we are developing to facilitate human rights reporting and verification. *See* www.thewhistle.org.

[3] I draw on my ongoing digital ethnography of human rights practices in the digital age for examples of empirical realities.

[4] R. Mansell, "The Life and Times of the Information Society" (2010) 28(2) *Prometheus: Critical Studies in Innovation* 165–86 at 173.

[5] K. Nash, *The Political Sociology of Human Rights* (Cambridge: Cambridge University Press, 2015).

advocacy work designed to impel change through persuasion.[6] At each stage, we can think of successful communication as speaking to and being heard and understood by intended audiences.[7] In other words, it is about audibility (or its equivalent, visibility). Unsuccessful communication, in contrast, involves either audibility to *un*intended audiences or *in*audibility to intended audiences. Successful communication can also have unsuccessful outcomes for the actors involved, as when a message is audible to intended audiences but is misunderstood or turns out to be erroneous, or when a message is received and interpreted as the communicator intended but turns out to be deceptive.

The success of communication matters, of course, for human rights practitioners' ability to generate accountability for individual cases of human rights violations. It also matters for a value at the core of human rights: pluralism, or the successful communication of a variety of voices. Three types of pluralism are of concern in this chapter. The first is the pluralism of human rights actors vis-à-vis the state or non-state actors they wish to hold to account. The second is the pluralism of individual human rights actors within the human rights world, which, as with all worlds, has hierarchies corresponding to the distribution of power.[8] The third is the pluralism of access by the subjects and witnesses of violations to the mechanisms of human rights accountability, which, of course, cannot act on a violation without first hearing about it.[9]

The chapter begins by outlining how risk is entwined with communication in the digital age. Rather than considering risk in isolation, we can think of it as manifesting via "risk assemblages," or dynamic combinations of actors, technologies, contexts, resources, and risk perceptions.[10] In the subsequent two sections, I detail selected types of risk for human rights communication resulting from new combinations of actors and technologies involved in digital fact-finding and advocacy. For fact-finding, these include the risk of surveillance, which has consequences for participants' physical security, and the risk of deception, which has consequences for their reputational integrity. For advocacy, these include the risk of mistakes,

[6] Human rights fact-finding is also used to produce evidence for courts, where the uptake of ICTs is also an important area for inquiry, but beyond the scope of this chapter.

[7] M. Madianou, L. Longboan, and J. C. Ong, "Finding a Voice through Humanitarian Technologies? Communication Technologies and Participation in Disaster Recovery" (2015) 9 *International Journal of Communication* 3020–38 at 3022; A. T. Thrall, D. Stecula, and D. Sweet, "May We Have Your Attention Please? Human-Rights NGOs and the Problem of Global Communication" (2014) 19(2) *The International Journal of Press/Politics* 135–59 at 137–38.

[8] W. Bottero and N. Crossley, "Worlds, Fields and Networks: Becker, Bourdieu and the Structures of Social Relations" (2011) 5(1) *Cultural Sociology* 99–119 at 105.

[9] E. McPherson, "Source Credibility as 'Information Subsidy': Strategies for Successful NGO Journalism at Mexican Human Rights NGOs" (2016) 15(3) *Journal of Human Rights* 330–46 at 331–32.

[10] D. Lupton, "Digital Risk Society," in A. Burgess, A. Alemanno, and J. O. Zinn (eds.), *The Routledge Handbook of Risk Studies* (Abingdon, UK: Routledge, 2016), p. 302.

which can in turn risk reputational integrity, and the risk of miscalculations, which can jeopardize precious resources. In the following section, I explain how this materialized risk combines with risk perceptions to create a silencing double bind. Human rights practitioners may be silenced if they don't know about risk – and they may silence themselves if they do. This silencing effect is not universal, however, but disproportionately affects human rights practitioners situated in more precarious contexts and with less access to resources.[11] This has consequences for the three types of pluralism outlined above. The chapter finishes by outlining four ways of loosening the risk double bind: educational, technological, reflexive, and discursive approaches to working with risk.

II COMMUNICATION, MEDIATION, AND RISK IN THE DIGITAL AGE

As communicators, we all do a number of things to increase the odds that our communications are successful. We establish the identities of our communication partners through clues we gather from their appearance and bearing. We supplement our messages with cues such as facial expressions or emoticons to guide interpretation, and we look for cues from our audiences that they have heard and understood us.[12] We gather information about our interlocutors' context – the time and place in which they are communicating – and supplement our messages with information about our own contexts (often referred to as metadata). We adjust our production and reception of content to these clues, cues, and contextual information. Still, even with all of these aids, communication can be unsuccessful, and this risk is exacerbated by the mediation of communication over ICTs.

Mediation is the extension of communication across time and/or space using technology. The closer we are in time and space to our communication partners, the easier it tends to be for us to establish their identities, observe and provide cues, and understand context. Easiest of all is face-to-face communication. By introducing "temporal and spatial distances," mediation makes all of this more difficult, as we are no longer in the same environment.[13] It is not, however, just this distance that increases the risk of unsuccessful communication, but also the introduction of intermediaries. These intermediaries are not neutral, but rather introduce new technical features and new actors with new motives, as well as new ways for existing actors to intervene with communications.

The technical features of ICTs can diminish, augment, or alter the clues, cues, and contextual metadata associated with a communication. Furthermore, the complexity of these technical features may make it difficult to understand just what has

[11] U. Beck, *Risk Society: Towards a New Modernity* (London: SAGE Publications, 1992), p. 23.
[12] J. B. Thompson, *The Media and Modernity: A Social Theory of the Media* (Stanford, CA: Stanford University Press, 1995), pp. 83–85.
[13] Ibid., p. 22.

happened. For example, many social media user profiles allow people to communicate with pseudonyms or assumed identities. Twitter's character limit squeezes nuance out of tweets, though users have introduced the use of hashtags as an abbreviated interpretation cue. In another example, YouTube automatically dates videos according to the day it is in California at the time of upload, no matter where the upload took place. This metadata is widely misunderstood and has contributed to disputes about the veracity of videos.[14]

A significant proportion of new actors behind ICTs are commercial, governed by profit motives. These motives can shape technical features, like Facebook's "like" and "share" buttons, which are designed to keep eyeballs on timelines peppered with advertisements. The motives of commercial communication platforms may not necessarily align with the motives of communicators. As discussed further below, this is particularly evident in the algorithms that determine visibility on social media and thus who is seen by whom. These algorithms may make certain communications either more or less visible than their producers intended. The phenomenon of commercial intermediaries controlling public visibility is nothing new – think of the gatekeeping role of mainstream news organizations. What is new is the lack of transparency and accountability when visibility decisions are made by a black-box algorithm instead of a human journalist.[15] Just as the technical complexity of ICTs obscures these algorithms and the commercial motives underpinning them, it also hides third-party actors. These include political actors who have a vested interest in human rights communication. The market for digital surveillance is thriving, and hardware and software that allow us to communicate over time and space also create opportunities for eavesdropping. In sum, at the same time as communicators using ICTs usually can glean less about their interlocutors and eavesdroppers than in a face-to-face situation, they must also know more about intermediary technologies that are both complex and opaque.[16] Mediation thereby increases the risk of unsuccessful communication and its attendant consequences.

Alongside many other professional worlds of communication, human rights practitioners are considering and adopting new ICTs. This use of ICTs and the mediation they engender supplements other forms of communication, creating a new "interaction mix" characterized by renewal, as fresh technologies proliferate and slightly stale ones become obsolete.[17] In terms of human rights fact-finding, this

[14] R. Mackey, "Confused by How YouTube Assigns Dates, Russians Cite False Claim on Syria Videos," *The New York Times*, August 23, 2013, http://thelede.blogs.nytimes.com/2013/08/23/confused-by-how-youtube-assigns-dates-russians-cite-false-claim-on-syria-videos/.

[15] Z. Tufekci, "Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency" (2015) 13 *Journal on Telecommunications and High Technology Law*: 203–18 at 208–09.

[16] E. McPherson, "Social Media and Human Rights Advocacy," in H. Tumber and S. Waisbord (eds.), *The Routledge Companion to Media and Human Rights* (London: Routledge, 2017), pp. 281–83.

[17] Thompson, *The Media and Modernity*, p. 87.

new mix has been described as enabling a new generation of methodologies.[18] Traditionally, the gold standard of fact-finding has been the face-to-face interview between civilian witnesses and human rights practitioners. Often facilitated by trusted networks cultivated over time, the interview allows for the co-production of information between the witness and the practitioner. This witness testimony and the accompanying analysis done by human rights practitioners are the cornerstones of the weighty, precisely worded, and highly documented orthodox human rights report.[19] These reports, in turn, underpin human rights advocacy, which practitioners traditionally – though not exclusively – communicated to targets via the mainstream media.[20]

Human rights communication has therefore always been mediated, whether information is passed through a trusted network of witnesses or shaped to attract the attention of journalists covering human rights violations.[21] It has also always entailed risk. One only has to dip into the multitude of reports on the conditions of human rights practice to see this – or to consider practitioners' risk-mitigation tactics, ranging from security training to robust and transparent methodologies to publicity strategies.[22] But the new mix of human rights fact-finding and advocacy in the digital age has brought about new risk assemblages shaped by technologies, actors, contexts, resources, and risk perceptions.[23] Over the next three sections of this chapter, I outline elements of these new risk assemblages and explain how they can hinder successful communication, with implications for the pluralism of human rights communication.

III DIGITAL FACT-FINDING AND COMMUNICATION RISK

Human rights practitioners have adopted ICTs for fact-finding in a variety of ways, including using high-technology information sources like satellite images, drone

[18] P. Alston, "Introduction: Third Generation Human Rights Fact-Finding," in *Proceedings of the Annual Meeting* (Washington, DC: American Society of International Law, 2013), pp. 61–62. For a recent overview of ways that ICTs are being adopted in human rights practice, *see* E. McPherson, *ICTs and Human Rights Practice* (Cambridge: University of Cambridge Centre of Governance and Human Rights, 2015).

[19] P. Alston and C. Gillespie, "Global Human Rights Monitoring, New Technologies, and the Politics of Information," *European Journal of International Law* (2012) 23(4) 1089–123 at 1108–09.

[20] M. Powers, "NGO Publicity and Reinforcing Path Dependencies: Explaining the Persistence of Media-Centered Publicity Strategies" (2016) 21(4) *The International Journal of Press/Politics* 492–94.

[21] McPherson, "Source Credibility as 'Information Subsidy'," at 333–35.

[22] S. Hopgood, *Keepers of the Flame: Understanding Amnesty International* (Ithaca, NY: Cornell University Press, 2006), pp. 90–92; A. M. Nah et al., "A Research Agenda for the Protection of Human Rights Defenders" (2013) 5(3) *Journal of Human Rights Practice* 401–20 at 413.

[23] S. Hankey and D. Ó Clunaigh, "Rethinking Risk and Security of Human Rights Defenders in the Digital Age" (2013) 5(3) *Journal of Human Rights Practice* 535–47 at 539.

videos, big data, and statistics as well as open source social media content.[24] Given our concern with communication, here I focus on practitioners' use of digital information that documents human rights violations and has been produced and transmitted by civilian witnesses – "civilian" in contrast with professional to highlight their inexpert status, and "witness" as someone who is purposively communicating experienced or observed suffering.[25] Civilian witnesses can be spontaneous or solicited.[26] In the digital age, spontaneous witnesses might use their smartphones to document violations that they then share with broader audiences via social media or messaging apps; sometimes this information is gathered, curated, and connected to human rights NGOs by networks of activists. Solicited witnesses may be answering a human rights NGO's open call for information made via a digital crowdsourcing project or a digital reporting application.

Digital information from civilian witnesses affords human rights practitioners a number of fact-finding advantages. First, the images and video civilian witnesses produce can provide much more detailed evidence than witness interviews that rely on memory.[27] Second, consulting civilian witnesses can tap wells of knowledge, particularly expertise relating to local contexts unfamiliar to foreign practitioners. Third, a wider incorporation of civilians via ICTs can fire up public enthusiasm about human rights and thus receptivity to advocacy.[28] Fourth, and most important for our concern with pluralism, these new sources can support the variety and volume of voices speaking and being heard on human rights. They supplement interviewing's traditional co-production of information between witnesses and practitioners with both the more autonomous production of spontaneous digital witnesses and new forms of co-production via solicited digital witnesses.[29] If these

[24] Amnesty International, Benetech, and The Engine Room, *DatNav: New Guide to Navigate and Integrate Digital Data in Human Rights Research* (London: The Engine Room, 2016). See also M. Latonero, "Big Data Analytics and Human Rights: Privacy Considerations in Context," Chapter 7 in this volume. Open source social media content includes perpetrator propaganda videos. It also includes content originally posted without a witnessing purpose but later repurposed by others, such as the use of a geolocated selfie for corroboration of a military vehicle's movements because it happens to capture that vehicle driving past in the background.

[25] E. McPherson, "Digital Human Rights Reporting by Civilian Witnesses: Surmounting the Verification Barrier," in R. A. Lind (ed.), *Produsing Theory in a Digital World 2.0: The Intersection of Audiences and Production in Contemporary Theory* (New York: Peter Lang Publishing, 2015), vol. 2, p. 206; S. Tait, "Bearing Witness, Journalism and Moral Responsibility" (2011) 33(8) *Media, Culture & Society* 1220–35 at 1221–22.

[26] McPherson, *ICTs and Human Rights Practice*, pp. 14–17.

[27] C. Koettl, *Citizen Media Research and Verification: An Analytical Framework for Human Rights Practitioners* (Cambridge: University of Cambridge Centre of Governance and Human Rights, 2016), p. 7.

[28] M. Land et al., *#ICT4HR: Information and Communication Technologies for Human Rights* (Washington, DC: The World Bank Group, 2012), p. 17; M. Land, "Peer Producing Human Rights" (2009) 46(4) *Alberta Law Review* 1115–39 at 1120–22.

[29] J. Aronson, "The Utility of User-Generated Content in Human Rights Investigations," Chapter 6 in this volume.

witnesses are situated in closed-country contexts or rapidly unfolding events, they might otherwise be inaccessible to human rights practitioners.[30] Indeed, fact-finding in a number of recent cases has hinged on evidence documented digitally by civilian witnesses. For example, Amnesty International's research into a 2017 shooting at an Australian refugee detention center in Papua New Guinea used refugees' photos and videos to challenge both governments' official version of events, which was that Papua New Guinea Defence Force soldiers fired bullets into the air rather than into the center.[31]

These opportunities are all made possible by ICTs' mediation of communication over time and place. Of course, this mediation, and the intermediaries it requires, also introduces risk. Below, I outline two possible manifestations of communication risk and their consequences arising from the introduction of new technologies into fact-finding, associated new commercial actors, and new opportunities for existing actors to interfere with communications. The first is the risk of surveillance, in which the communication is audible to unintended recipients and generates concomitant risk for the physical security of civilian witnesses and human rights practitioners. The second is the risk of deception, in which the producer of a digital communication engineers the recipient's misinterpretation of that communication. Misinterpretation creates follow-on risk to the reputational integrity of human rights practitioners and their NGOs. These are familiar categories of risk in the human rights domain but manifested, as explained below, in new ways. Both are made possible by the technical complexity of mediating ICTs, which allows eavesdroppers to hide and deceivers to manipulate metadata.

A *Surveillance and Physical Security*

Surveillance, understood broadly as monitoring information about others for purposes including management and control, is a risk that civilian witnesses and human rights practitioners have always faced.[32] Surveillance of their identities, networks, and activities is a key tactic deployed by state adversaries in a "cat-and-mouse" game over truth-claims.[33] Human rights practitioners who pioneered the use of ICTs may have had a momentary advantage in this battle by using these technologies to transmit information quickly and widely. Many state actors, however, have caught up quickly and even surpassed human rights actors in their strategic use of ICTs.

[30] Alston and Gillespie, "Global Human Rights Monitoring, New Technologies, and the Politics of Information," at 1112–13.

[31] "In the Firing Line: Shooting at Australia's Refugee Centre on Manus Island in Papua New Guinea," Amnesty International, May 14, 2017 www.amnesty.org/en/documents/document/?indexNumber=asa34%2f6171%2f2017&language=en.

[32] D. Lyon, *Surveillance after Snowden* (Cambridge: Polity Press, 2015), p. 3.

[33] Hankey and Ó Clunaigh, "Rethinking Risk and Security of Human Rights Defenders in the Digital Age," at 538.

The surveillance opportunities ICTs afford center on a metadata paradox. ICTs can both reveal and conceal communication metadata; the first facilitates mass surveillance, while the second facilitates spyware.

ICTs are built to collect metadata on their users, often without users understanding just how significant their data trails are. Many ICT companies routinely collect users' metadata for reasons ranging from marketing to legal compliance.[34] This profit-driven surveillance produces information about communications that also meets the surveillance imperatives of states. The US National Security Agency, for example, infamously has a bulk surveillance program that collects telecommunications metadata. Activists worry that this program has set a standard for other governments in terms of the permissible level of spying on their citizenries – as exemplified by the Egyptian government's 2014 request for tenders for a mass social media surveillance system.[35] In addition to its implications for the rights to privacy and freedom of opinion and expression, this form of surveillance is a particular concern for individuals communicating information critical of retaliatory states.[36] Even if the content of these communications remains private, metadata can reveal connections between civilian witnesses and human rights practitioners and, through social network analysis, identify individuals as human rights practitioners.[37]

While mass surveillance depends on ICTs' revelation of communication metadata, spyware depends on its obfuscation, afforded by ICTs' complexity. Spyware hides in victims' communications equipment to track and share information about their activities.[38] In order to get spyware into target devices in the first place, victims must be deceived into installing it. This often happens through a wolf-in-sheep's-clothing tactic called social engineering, where messages containing spyware are disguised through the manipulation of metadata and content. For example, a human rights practitioner in the United Arab Emirates received unsolicited text messages containing a link that appeared to document evidence of prison torture. Had he clicked on the link, this practitioner's iPhone would have been infected with commercial spyware priced at around $1 million – an indication that a powerful

[34] "Metadata," Privacy International, www.privacyinternational.org/node/53.

[35] "Egypt's plan for mass surveillance of social media an attack on internet privacy and freedom of expression," Amnesty International, June 4, 2014, www.amnesty.org/en/latest/news/2014/06/egypt-s-attack-internet-privacy-tightens-noose-freedom-expression/; S. Kelly et al., "Tightening the Net: Governments Expand Online Controls," Freedom House, 2014, https://freedomhouse.org/report/freedom-net/2014/tightening-net-governments.

[36] *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye*, U.N. Doc. A/HRC/29/32 (May 22, 2015).

[37] Amnesty International, Benetech, and The Engine Room, *DatNav*, p. 23; S. Bender-de Moll, *Potential Human Rights Uses of Network Analysis and Mapping* (Washington, DC: AAAS Science and Human Rights Program, 2008), p. 4.

[38] M. Schwartz, "Cyberwar for Sale," *The New York Times*, January 4, 2017, www.nytimes.com/2017/01/04/magazine/cyberwar-for-sale.html.

actor was behind the attack.[39] In another case, a Mexican human rights practitioner received a text message purporting to share news about the investigation into the 2014 disappearance of forty-three students. He fell for it, clicking on the link and infecting his phone with malware believed to have been sold to the Mexican government by an Israeli cyberwarfare company.[40]

Digital security risk turning into physical security risk is unfortunately becoming more and more common for human rights practitioners and civilian witnesses.[41] If surveillance makes fact-finding communication audible to an unintended audience, its participants may not be aware this has happened until they experience related harassment and attacks. Security risk may spread through practitioners' and witnesses' networks, which are rendered visible by smartphone contacts and social media friends and followers lists. Furthermore, the mediation of digital fact-finding over time and space can make it difficult for practitioners who have learned of a threat to locate and warn civilian witnesses.[42] Human rights practitioners can and do use security tools – such as technologies supporting encryption, anonymity, and the detection of spyware – to counteract the corporate/state surveillance nexus. These technologies are threatened, however, by laws curtailing their use.[43] Furthermore, powerful discourses, such as "nothing to hide, nothing to fear," which have been propagated by state actors and picked up by the media, align the use of these technologies with criminality and threats to national security.[44]

B *Deception and Reputational Integrity*

Human rights practitioners' use of digital information from civilian witnesses generates another category of risk: susceptibility to misinterpretation through deception. By dint of their accusations of violations, human rights practitioners often engage in battles over truth-claims with their adversaries. Though the manipulation of truth-claims with an intent to deceive has always been a feature of these battles, human rights practitioners may be more exposed to them in the digital age for several reasons. First, ICTs afford a greater number and variety of sources of information,

[39] B. Marczak and J. Scott-Railton, "The Million Dollar Dissident: NSO Group's iPhone Zero-Days Used against a UAE Human Rights Defender," The Citizen Lab, August 24, 2016, https://citizenlab.org/2016/08/million-dollar-dissident-iphone-zero-day-nso-group-uae/.

[40] A. Ahmed and N. Perlroth, "Using Texts as Lures, Government Spyware Targets Mexican Journalists and Their Families," *The New York Times*, June 19, 2017, www.nytimes.com/2017/06/19/world/americas/mexico-spyware-anticrime.html?_r=0.

[41] S. Kelly et al., "Silencing the Messenger: Communication Apps Under Pressure," Freedom House, 2016, https://freedomhouse.org/report/freedom-net/freedom-net-2016.

[42] Amnesty International, Benetech, and The Engine Room, *DatNav*, p. 61.

[43] A. Crowe, S. Lee, and M. Verstraete, "Securing Safe Spaces Online: Encryption, Anonymity, and Human Rights," Privacy International, 2015, www.privacyinternational.org/sites/default/files/Securing%20Safe%20Spaces%20Online_0.pdf.

[44] H. Abelson et al., "Keys under Doormats: Mandating Insecurity by Requiring Government Access to All Data and Communications," (2015) 1(1) *Journal of Cybersecurity* 69–79.

many of whom are outside of the trusted networks that human rights organizations traditionally consult. Deceptive actors can camouflage themselves among this broader pool of sources. Second, unlike in a traditional face-to-face interview, human rights practitioners using spontaneous or solicited digital information from civilian witnesses are not present at the moment of production. As such, they cannot rely on their direct perceptions of identity clues, communication cues, and contexts to verify civilian witnesses' accounts.[45] Instead, they must use digitally mediated content and metadata as a starting point, which can be distorted and manipulated. Third, this information is often in image or video format that appears to be amateur. This lends it an aura of authenticity – rooted, perhaps, in a "seeing is believing" epistemology – that may belie manipulation.[46]

Deception through truth-claims manipulation can be divided into at least three categories: outright staging of content, doctoring of content, and doctoring of metadata.[47] Staging of content involves packaging fakery as fact, as with the viral YouTube video "SYRIA! SYRIAN HERO BOY rescue girl in shootout." This video, which claimed to document children dodging bullets while running through a dusty Syrian street, was actually a cinematographic project by a Norwegian director that was filmed in Malta.[48] Doctored content, in turn, uses real rather than staged content but relies on digital editing tools such as Photoshop to alter the images. For example, one human rights practitioner received images via WhatsApp from a source who claimed that they were evidence of torture during detention. These included a picture of a person who seemed, at first glance, to have a bruised face. Additional investigation, however, revealed that this was a highly edited version of an older picture involving changes to its color balance to create the illusion of bruises.[49]

Human rights practitioners report that it is the last of these three forms of deception – the doctoring of metadata – that is by far the most prevalent.[50] This involves scraping videos or images from one context and repackaging them as evidence of violations in another context. Examples include reposting YouTube videos with new descriptions, as in the case of one video depicting the water cannoning of a man shackled to a tree while other men watch and laugh. This video appeared on YouTube multiple times with at least three different sets of metadata entered in the video description. One version claimed to depict Venezuelan armed forces assailing a student, another stated that it was Colombian

[45] Diane F. Orentlicher, "Bearing Witness: The Art and Science of Human Rights Fact-Finding" (1990) 3 *Harvard Human Rights Journal* 83–136 at 114.

[46] P. Brown, *"It's Genuine, as Opposed to Manufactured": A Study of UK News Audiences' Attitudes towards Eyewitness Media* (Oxford: Reuters Institute for the Study of Journalism, 2015), http://reutersinstitute.politics.ox.ac.uk/publication/its-genuine-opposed-manufactured.

[47] Amnesty International, Benetech, and The Engine Room, *DatNav*, p. 35.

[48] McPherson, "Digital Human Rights Reporting by Civilian Witnesses," pp. 193–94.

[49] Koettl, *Citizen Media Research and Verification*, pp. 27–28.

[50] Ibid., p. 16.

https://doi.org/10.1017/9781316838952.009 Published online by Cambridge University Press

special forces and a farmer, and a third portrayed the scene as Mexican police and a member of a civil defense group.[51]

Though some instances of deception may be malevolent, other instances may be backed by the best of intentions. For example, civilian witnesses may use images from one event to illustrate another, similar event that was not recorded. Nevertheless, using any kind of manipulated information as evidence creates a follow-on risk to the reputations of human rights practitioners and their organizations. For these, credibility is a fundamental asset, not only for the persuasiveness of their advocacy, but also for garnering donations and volunteers, influencing policy-making, and motivating mobilization.[52] Credibility is also a human rights organization's Achilles' heel, as it can be damaged in an instant with the publication of truth-claims that others convincingly expose as false.[53] Though the verification of information has always been a cornerstone of human rights work as a truth-claim profession, information mediated by ICTs is challenging established verification practices. This is not only because of the new sources and formats of information ICTs enable, but also because verifying digital information requires expertise that, though increasingly standardized, is still emergent.[54]

IV DIGITAL ADVOCACY AND COMMUNICATION RISK

As with fact-finding, human rights practitioners are incorporating ICTs into their advocacy strategies, venturing far beyond websites into formats including apps, livestreaming, and virtual reality. Because human rights practitioners are paying particular attention to mainstream social media platforms to supplement their traditional advocacy practices, I focus on that medium here.[55] Practitioners are communicating advocacy messages via social media to directly target policy-makers, either publicly or via private messages, and to attract the attention of the mainstream media.[56] They are also using social media to mobilize publics for a variety of reasons, including fundraising, creating visibility for an issue, and building networks

[51] M. Bair and V. Maglio, "Video Exposes Police Abuse in Venezuela (Or Is It Mexico? Or Colombia?)," *WITNESS Blog*, February 25, 2014, http://blog.witness.org/2014/02/video-exposes-police-abuse-venezuela-mexico-colombia/.

[52] L. D. Brown, *Creating Credibility* (Sterling, VA: Kumarian Press, 2008), pp. 3–8; S. Cottle and D. Nolan, "Global Humanitarianism and the Changing Aid-Media Field: Everyone Was Dying for Footage" (2007) 8(6) *Journalism Studies* 862–88 at 872; M. Gibelman and S. R. Gelman, "A Loss of Credibility: Patterns of Wrongdoing Among Nongovernmental Organizations" (2004) 15(4) *Voluntas: International Journal of Voluntary and Nonprofit Organizations* 35–81 at 372.

[53] Koettl, *Citizen Media Research and Verification*, p. 6.

[54] McPherson, "Digital Human Rights Reporting by Civilian Witnesses," pp. 199–200.

[55] "Incorporating Social Media into Your Human Rights Campaigning," New Tactics in Human Rights, 2013, www.newtactics.org/conversation/incorporating-social-media-your-human-rights-campaigning.

[56] Powers, "NGO Publicity and Reinforcing Path Dependencies," p. 500.

between publics and subjects of violations in a show of global solidarity.[57] Many NGOs undertake advocacy over social media through institutional accounts operated by individuals. Though dedicated communications professionals operate these accounts at some organizations, at others the arrangement is more ad hoc, undertaken by existing staff according to interest or availability.

The use of social media affords human rights practitioners a number of advocacy advantages. It can allow them to amplify messages and reach advocacy targets without depending on the mainstream media, whose human rights coverage may be circumscribed by commercial imperatives, censorship, and norms of newsworthiness.[58] The range of communication formats supported by social media enables development of new and captivating ways to represent human rights information, such as data visualization.[59] Additionally, the quantification metrics built into social media platforms, such as numbers of likes and shares, allow human rights practitioners to track engagement with their messages.[60] They can then incorporate these numbers into their campaigns targeted at policy-makers to quantify public support for their advocacy aims.[61] A wide variety of human rights advocacy communications over social media exists, such as the 2013 Thunderclap campaign created by EDUCA, an NGO based in Oaxaca, Mexico, to raise awareness about ongoing human rights violations there. Thunderclap is a digital platform that allows users to coordinate their supporters' automatic participation in a onetime, synchronized mass social media posting of a particular message.[62] EDUCA surpassed its goal of 100 supporters, and its Thunderclap – timed to coincide with the October 23 UN Universal Periodic Review of Mexico's human rights record – reached more than 58,000 people via social media.[63]

[57] McPherson, *ICTs and Human Rights Practice*, pp. 28–32; R. Stewart, "Amnesty International's head of comms on why interactive social campaigns could help find a solution to the refugee crisis," *The Drum*, February 7, 2017, www.thedrum.com/news/2017/02/07/amnesty-international-s-head-comms-why-interactive-social-campaigns-could-help-find.

[58] Alston and Gillespie, "Global Human Rights Monitoring, New Technologies, and the Politics of Information," pp. 1112–13; E. McPherson, "How Editors Choose Which Human Rights News to Cover: A Case Study of Mexican Newspapers," in T. A. Borer (ed.), *Media, Mobilization, and Human Rights: Mediating Suffering* (London: Zed Books, 2012), pp. 96–121.

[59] J. Emerson et al., "The Challenging Power of Data Visualization for Human Rights Advocacy," Chapter 8 in this volume.

[60] D. Karpf, *The MoveOn Effect: The Unexpected Transformation of American Political Advocacy* (New York: Oxford University Press, 2012), pp. 36–37.

[61] E. McPherson, "Advocacy Organizations' Evaluation of Social Media Information for NGO Journalism: The Evidence and Engagement Models" (2015) 59(1) *American Behavioral Scientist* 124–48 at 134–39.

[62] Thunderclap is free, but the platform does decide whether or not to approve campaigns, and the extent of campaign visibility can depend on users' purchase of premium plans. "Take your message even further," Thunderclap, 2017, www.thunderclap.it/pricing.

[63] EDUCA, "Thunderclap: TÚ PUEDES EVALUAR A EPN EN DH," October 23, 2013, www.thunderclap.it/projects/5687-t-puedes-evaluar-a-epn-en-dh.

Again, however, the advantages of social media for advocacy are accompanied by risk, and here I detail two types of communication risk and their consequences. Both stem from the introduction of new technologies into advocacy, which in turn introduces new actors with new motives. Human rights practitioners are accustomed to considering the motives of intermediaries and their intended audiences in shaping their advocacy strategies. For example, they cater to the "media logic" of mainstream media outlets, tailoring the tone and theme of their content as well as building their identities as credible sources to meet journalists' exigencies.[64] The use of social media intermediaries, however, requires them to shape advocacy messages in light of new "social media logics."[65] These are also commercially driven motives, manifested in new technical features. Like journalists and journalism, these technical features can be inscrutable to human rights practitioners and incompatible with human rights advocacy – but in different ways.[66] Conducting advocacy via these intermediaries thus introduces new facets to existing risk. This risk includes audibility to *un*intended audiences, which I refer to as mistakes that can have reputational consequences. The second variety of risk addressed below is *in*audibility to intended audiences, or advocacy miscalculations that waste resources.

A *Mistakes and Reputational Integrity*

An advocacy-related mistake involves something happening that the communication's producer does not wish to happen. Social media's facilitation of mediation to publics, in combination with technical features that both speed up and obscure the dynamics of this mediation, introduce new ways of making mistakes. Analog means of communicating with publics had areas of friction, such as the effort required to set up a press conference.[67] This friction, no doubt, was frustrating during crises in need of immediate response, but it also allowed room for reflexivity and proofing. Digital communication to publics, in contrast, requires only the click of a button. As such, the pace of communication is much faster on social media, as is the pressure to produce at speed. Proofing becomes the friction, and there may not always be time for this to be done as thoroughly as one would like. Furthermore, the technical complexity of social media can make proofing difficult to do. This is particularly the case with respect to ensuring that the right communication is audible to the right audience, as audiences are both blurred and obscured by social media. Mistakes can

[64] S. Waisbord, "Can NGOs Change the News?" (2011) 5 *International Journal of Communication* 142–65 at 149–51.

[65] J. van Dijck and T. Poell, "Understanding Social Media Logic" (2013) 1(1) *Media and Communication* 2–14.

[66] McPherson, "Social Media and Human Rights Advocacy."

[67] S. Gregory, "Human Rights Made Visible: New Dimensions to Anonymity, Consent, and Intentionality," in M. McLagan and Y. McKee (eds.) *Sensible Politics: The Visual Culture of Nongovernmental Activism* (New York, Cambridge, MA: Zone Books, 2012), p. 552.

thus be about erroneous content, but they can also involve the transmission of private information to publics or of information intended for one "imagined audience" or communication context to another.[68] The consequences of these mistakes are also caught up with mediation, as ICTs allow endless possibilities of repetition and amplification over time and space.[69]

Here, I develop the example of individual practitioners managing multiple Twitter accounts, each with its own profile and audience. Rather than involving an error in the advocacy itself, the associated mistake results from having social media open as an advocacy channel. Twitter's phone apps easily allow users to switch between accounts, requiring nothing more than holding down the profile icon in the iPhone version. Of course, this also means it is easy to slip between accounts erroneously or forgetfully. When one account is personal and one is institutional, this can create some sticky situations.

One such situation arose in response to a 2014 tweet by Amnesty International about the police shooting in Ferguson, Missouri: "US can't tell other countries to improve their records on policing and peaceful assembly if it won't clean up its own human rights record." Six minutes later, the Center for Strategic and International Studies (CSIS), a major public policy think tank, replied, "Your work has saved far fewer lives than American interventions. So, suck it." CSIS scrambled to quickly explain the tweet as the work of an intern who had access to the CSIS Twitter account but thought he was logged into his personal account instead when he wrote the message. In the context of a flurry of media stories, CSIS's senior vice president of external relations described himself and his colleagues as "distressed," and CSIS quickly sent out an apology tweet to Amnesty. Amnesty followed this by tweeting: ".@CSIS and @amnesty have kissed and made up. Now back to defending human rights!"[70]

Though this example is relatively lighthearted, more serious mistakes can have more serious consequences. One human rights practitioner told me about a mistake made on his organization's Facebook feed when an image of a private meeting was erroneously published. A furious phone call from an important participating organization ensued, creating what the practitioner described as "a terror effect within the organization" about using social media. At the time of our interview, the resulting policy at this NGO was that every social media post made on the institutional account must first be approved by the executive director.

[68] A. E. Marwick and D. Boyd, "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience" (2011) 13(1) *New Media & Society* 114–33; Thompson, *The Media and Modernity*, pp. 143–44.

[69] Thompson, *The Media and Modernity*, p. 141.

[70] B. James, "Think Tank Apologizes for Intern's 'Suck It' Tweet to Amnesty International," *Talking Points Memo*, August 19, 2014, http://talkingpointsmemo.com/livewire/csis-amnesty-international-suck-it-tweet; M. Roth, "Think Tank Blames Intern for Tweet Telling Amnesty International to 'Suck It'," *MTV News*, August 20, 2014, www.mtv.com/news/1904747/csis-intern-amnesty-international/.

Serious mistakes can jeopardize an organization's reputational integrity, particularly with respect to credibility and professionalism. The relative permanence of information published on social media, as well as the unpredictability of its circulation, means a mistake cannot be undone but must instead be overcome. Repairing a damaged reputation, which may involve performing credibility over time and rebuilding social capital, can divert precious resources from human rights NGOs' core aims.[71] Even if the mistake is quickly forgiven, it can – as Amnesty's last tweet above highlights – detract from the message and work of the organization. Because of the risk of mistakes that accompanies the use of social media, adopting this technology can result in slower and more resource-intensive practices than expectations might suggest.

B *Miscalculation and Resources*

A communication miscalculation means that one's message is inaudible to one's intended audience. Of course, the risk always exists that one's audience either does not hear or does not listen to the message. This is exacerbated by mediation, not only because distance makes it more difficult to perceive audience cues about attention, but also because the intermediary may do things to the message to make it less audible. In the case of social media, this includes evaluating messages automatically with timeline algorithms to determine how visible they should be, and to whom.

Human rights practitioners are in good company with respect to not knowing exactly how these algorithms make decisions about message visibility. The algorithms that govern social media timeline visibility are considered proprietary trade secrets, and these algorithms in turn may be governed by deep learning, in which the algorithm adapts autonomously based on the information to which it is applied.[72] Furthermore, these algorithms may have thousands of moving parts that are updated weekly or even daily.[73] Deciphering these algorithms – which are black boxes to just about everybody, even possibly to those who design them – is a far cry from building a trusting relationship with a journalist.[74]

Practitioners do know that these algorithms prevent organizations from reaching all of their fans or followers with their posts. "Organic," or unpaid, reach may be only 10 percent of a potential audience, and only a small proportion of those reached

[71] Cottle and Nolan, "Global Humanitarianism and the Changing Aid-Media Field" at 871–74.

[72] N. Koumchatzky and A. Andryeyev, "Using Deep Learning at Scale in Twitter's Timelines," Twitter, May 9, 2017, https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.html; Tufekci, "Algorithmic Harms Beyond Facebook and Google."

[73] W. Oremus, "Twitter's New Order," *Slate*, March 5, 2017, www.slate.com/articles/technology/cover_story/2017/03/twitter_s_timeline_algorithm_and_its_effect_on_us_explained.html.

[74] W. Knight, "The Dark Secret at the Heart of AI," *MIT Technology Review*, April 11, 2017, www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/; McPherson, "Source Credibility as 'Information Subsidy'."

will engage with the post by liking, sharing, or clicking on a link.[75] Facebook does shed some light on how this organic reach is determined, stating in its support materials for nonprofits that the post's timing and its relevance to particular audience members matter.[76] Twitter reveals that it ranks a tweet for relevance on a number of criteria, including how much user interaction it has already generated and how much past interaction exists between the producer and the potential recipient; in other words, visibility returns to the already visible and to the already networked.[77] Still, ambiguity remains for the organic visibility of individual posts. Greater certainty is available, however – at a price: mainstream social media platforms allow users to buy access to larger and targeted audiences. Social media advocacy is therefore a "free-to-play, pay-to-win game."[78]

Human rights practitioners encounter further elements of social media logic that generate communication risk. One is social media platforms' community standards, which outline the grounds for removal of content that might alienate users. Graphic images and videos fall into this category. The problem for human rights advocacy as well as fact-finding is that the documentation of certain categories of violations necessarily involves depictions of violence – though practitioners think through the ethics of such representations very carefully.[79] Like the determination of timeline visibility, content moderation is an opaque decision-making process.[80] Practitioners know that whether or not a graphic video or image stays on social media can depend on a number of factors, including how it is explained by whoever posts it (Facebook allows graphic images and videos to stay up if they are "in the public interest," but not if they are "for sadistic pleasure"), if it is reported by another user, what the content moderator employed by the platform to review content decides, and – as recently happened with the video livestreamed by Diamond Reynolds immediately after police shot her boyfriend – even "technical glitches."[81]

[75] M. Collins, "It's time for charities to stop wasting money on social media," *The Guardian*, March 11, 2016, www.theguardian.com/voluntary-sector-network/2016/mar/11/charities-wasting-money-social-media.

[76] Facebook, "Measurement & Tracking," Nonprofits on Facebook, 2017, https://nonprofits.fb.com/topic/measurement-tracking/.

[77] Koumchatzky and Andryeyev, "Using Deep Learning at Scale in Twitter's Timelines."

[78] L. Karch, "Is Social Media a Time-Waster for Nonprofits?" *Nonprofit Quarterly*, March 17, 2016, https://nonprofitquarterly.org/2016/03/17/is-social-media-a-time-waster-for-nonprofits/.

[79] M. Bair, "Navigating the Ethics of Citizen Video: The Case of a Sexual Assault in Egypt" (2014) 19 *Arab Media & Society* 1–7; Gregory, "Human Rights Made Visible," p. 555.

[80] S. T. Roberts, "Commercial Content Moderation: Digital Laborers' Dirty Work," in S. Umoja Noble and B. M. Tynes (eds.), *The Intersectional Internet: Race, Sex, Class, and Culture Online* (New York: Peter Lang Publishing, 2016), pp. 148–49.

[81] "Community Standards," Facebook, 2017, www.facebook.com/communitystandards#violence-and-graphic-content; A. Peterson, "Why the Philando Castile police-shooting video disappeared from Facebook – then came back," *The Washington Post*, July 7, 2016, www.washingtonpost.com/news/the-switch/wp/2016/07/07/why-facebook-took-down-the-philando-castile-shooting-video-then-put-it-back-up/.

A third way in which social media logics can introduce advocacy miscalculations is the content culture they cultivate by rewarding certain types of content with visibility – a culture that contrasts sharply with the traditional registers of human rights advocacy.[82] Facebook, for example, counsels nonprofits that "formal language can feel out of place" and that "placing blame … typically doesn't lead to high engagement."[83] It may also be that certain types of human rights and certain types of victims are more aligned than others with the logics of social media virality, which is co-constructed by the predilections of algorithms and networked humans.[84] This was the topic of much public contemplation following the 2015 circulation of an image of three-year-old Syrian refugee Alan Kurdi's body washed up on a Turkish beach. Many critically attributed this image's viral spread to Alan's resemblance to a Western child, and thus his relatability to Western social media users.[85] Furthermore, the competition for audience attention on social media has fueled the rise of "clickbait" headlines, which feature a "curiosity gap." These headlines give away just enough to pique someone's attention, but require that person to click on a link to get the full story.[86] An interviewee from a human rights NGO that works with migrants and refugees joked about why this popular format is not an option for her organization's advocacy practices: "We are not going to be like, you know, 'This man got to the border, and you would never believe what happened next!' You can't do that, because it makes you sound … your credibility is gone. So we don't do that." The content culture that is rewarded on social media, then, may also be at odds with what the target audiences of human rights advocacy want to hear from practitioners – if the audience even pays attention to social media advocacy in the first place.[87]

Using social media allows human rights practitioners to directly address advocacy targets, but whether those targets hear or listen to those advocacy messages is often an open question. The risk of such advocacy miscalculation generates follow-on risks to an NGO's resources. These include wasted time, since maintaining a social media presence – including designing content, building and interacting with networks, and developing advertising strategies – demands significant person-hours. This is also a waste of money, as is targeted advertising that falls on deaf ears. Social media's relative novelty has meant a steep learning curve for human rights

[82] McPherson, "Social Media and Human Rights Advocacy," pp. 281–83.
[83] "Grab People's Attention," Nonprofits on Facebook, 2016, https://nonprofits.fb.com/topic/grab-peoples-attention.
[84] van Dijck and Poell, "Understanding Social Media Logic," p. 7.
[85] *See, e.g.*, C. Homans, "The Boy on the Beach," *The New York Times*, September 3, 2015, www.nytimes.com/2015/09/03/magazine/the-boy-on-the-beach.html.
[86] D. Thompson, "Upworthy: I Thought This Website Was Crazy, but What Happened Next Changed Everything," *The Atlantic*, November 14, 2013, www.theatlantic.com/business/archive/2013/11/upworthy-i-thought-this-website-was-crazy-but-what-happened-next-changed-everything/281472/.
[87] Powers, "NGO Publicity and Reinforcing Path Dependencies" at 498.

practitioners, and risk to advocacy communications can be diminished with expertise. At the same time, however, mastery remains somewhat of a mirage, due not only to the inaccessible element of social media logics, but also to the ICT sector's state of permanent renewal. Users regularly encounter new platforms as well as new features within the platforms they use, which appear seemingly overnight as tweaks to commercially driven systems designed to hold our attention.

So far, I have outlined the communication risk posed by digital fact-finding and advocacy related to new technologies and new actors; in the next section, I put these findings into conversation with contexts, resources, and risk discourses to show how risk's silencing effect is not universal, but rather can map onto existing inequalities.

V RISK ASSEMBLAGES, PLURALISM, AND INEQUALITY

Returning to the three types of pluralism introduced earlier in the chapter, it is clear that the manifested forms of risk outlined above have silencing effects on the first category – the pluralism of the human rights world vis-à-vis the world of power it aims to hold to account. New mediating technologies, with commercially driven technical features that complicate communication, fuel new communication cultures and allow new spaces for adversaries to intervene. Surveillance, through its consequences for physical security, can stop human rights practitioners from speaking. The susceptibility of practitioners to deception and mistakes, with the repercussions for reputations, may deafen advocacy targets to their communications. Advocacy miscalculations may prevent advocacy targets from hearing those communications at all. In order to understand the effects of communication risk on the other types of pluralism, however, we must further develop our understanding of these new risk assemblages. We must also think about context, resources, and risk discourses.

As materialized risks are always embodied, an individual practitioner's context and resources matter in understanding how risk impacts the second type of pluralism, namely pluralism within the human rights world. Context – or individuals' "social risk positions" ranging from their political environments to their positions within relevant social hierarchies – influences exposure to risk.[88] In turn, the resources individuals have at hand influence their ability to mitigate risk. Key here is the resource of expertise, such as digital literacy about computer security, knowledge of digital verification practices, and facility with social media. Also relevant are the resources that can be used to secure expertise, including money, of course, but also the social capital and reputations that can connect practitioners to expertise and convince experts to share it. The same resources can be used to secure physical and digital safeguards. The upshot is that risk curtails pluralism within the human rights world by silencing practitioners unequally.

[88] Beck, *Risk Society*, p. 23.

https://doi.org/10.1017/9781316838952.009 Published online by Cambridge University Press

Inequalities in contexts and resources intersect with the types of risk enumerated above in a variety of ways. The risk of surveillance depends greatly on the proclivities of a practitioner's political opponents for purchasing surveillance technologies and enacting pro-surveillance legislation. It also depends on the practitioner's networks, whose resistance to surveillance is only as strong as their weakest links; one member falling prey to malware can unwittingly expose all her communication partners.[89] Security literacy is crucial. As a practitioner at an organization that trains human rights reporters on digital security once told me, "A lot of them don't know that Facebook is where a lot of people who would target human rights defenders go shopping." Security literacy is expensive, in terms of money and time, and it is daunting; therefore, it is more accessible to some than to others.[90]

Deception via the manipulation of truth-claims is also a risk that human rights practitioners experience differently. Like surveillance, this risk is conditional on the political context, since some governments are particularly inclined to engage in information wars. Associated reputational risks are not isolated, but rather may have repercussions for a practitioner's networks. This is because human rights organizations build their credibility in part through networks of association with credible peers; one organization's loss of credibility allows opponents to tar its network with the same brush.[91] Some organizations can weather a hit on their credibility better than others. As human rights organizations also build credibility through performance over time, a more well-established NGO would have more reputational capital to counterbalance an instance of susceptibility to deception or a mediation-related mistake.[92]

The risk of advocacy mistakes and miscalculations can be mitigated by human rights organizations' in-house social media expertise and consequently the money required to acquire this expertise. Funds also allow human rights organizations to buy visibility for their social media communications through targeted advertisements. Those with fewer resources to dedicate to social media advocacy are, unfortunately, more likely to waste resources by engaging in this practice. This is evident in the results of a recent study, which found that, of 257 sampled human rights NGOs, the richest 10 percent had 92 percent of the group's total Twitter followers, 90 percent of their views on YouTube, and 81 percent of their likes on Facebook. The study also found that social media advocacy does not seem to help NGOs set the agenda in the mainstream media – further evidence that unsuccessful digital communication can curtail the greater pluralism that using ICTs could bring, both within the human rights world and vis-à-vis the world of power.[93]

[89] Kelly et al., "Tightening the Net."
[90] Hankey and Ó Clunaigh, "Rethinking Risk and Security of Human Rights Defenders in the Digital Age" at 542.
[91] M. Land, "Peer Producing Human Rights," at 1136; Gibelman and Gelman, "A Loss of Credibility" at 376.
[92] McPherson, "Source Credibility as 'Information Subsidy'" at 337.
[93] Thrall, Stecula, and Sweet, "May We Have Your Attention Please?" at 143.

A major purpose of the first and second forms of pluralism is to support the third form: the pluralism of civilian access to human rights mechanisms. Civilians cannot access accountability without their voices – their accounts – being heard. If the NGOs representing them are silenced, they too may be silenced. So, communication risk restricts civilian access to the mechanism of human rights unequally as well. As this effect maps onto context and resource distributions, this means that civilians in more precarious contexts with relatively few resources – in other words, those who might most need human rights mechanisms – are more likely to be silenced. The networked nature of human rights NGOs, which are characterized by solidarity, information exchange, and international communication, goes some way to counteract this effect, as another organization may be able to pick up the communication chain.[94] Still, while ICTs do create human rights communication channels where none existed before, we must be alert to the possibility that they do not level inequalities of audibility, but rather extend them.

So far, this chapter has looked at materialized risk, but risk perception is just as important for understanding the human rights practitioner's lived experience of risk.[95] It is also just as important for understanding how the risk accompanying use can impact the pluralizing potential of ICTs. As evident from interviews with some human rights practitioners, in which they qualified their view of ICTs with words such as "terrified" and "scary," knowing about risk can be distracting and even debilitating. The more complex the risk assemblage, the stronger this effect, as it is more difficult to understand and predict the risk. This knowing but not knowing *exactly* brings its own anxieties.[96]

Risk perception is not necessarily accurate, in part because risks are hard to estimate and because the idea of them can be overwhelming. Furthermore, as explained below, the specter of risk associated with a practice may have been conjured on purpose to prevent people from undertaking that practice; it may be a discourse deployed in the pursuit of power.[97] A full exploration of risk perception, which is outside the confines of this chapter, would consider the practices individuals adopt in anticipation of these risks and would investigate how these practices affect pluralism. For example, some human rights practitioners are renouncing digital communication methods for a return to analog.[98] Some are slow to adopt digital information from civilian witnesses for fact-finding.[99] As mentioned above,

[94] M. E. Keck and K. Sikkink, *Activists beyond Borders: Advocacy Networks in International Politics* (Ithaca, NY: Cornell University Press, 1998).

[95] Nah et al., "A Research Agenda for the Protection of Human Rights Defenders" at 405–06.

[96] Beck, *Risk Society*, pp. 22, 54.

[97] D. Lupton, "Introduction: Risk and Sociocultural Theory," in D. Lupton (ed.), *Risk and Sociocultural Theory: New Directions and Perspectives* (Cambridge: Cambridge University Press, 1999), pp. 4–5.

[98] Hankey and Ó Clunaigh, "Rethinking Risk and Security of Human Rights Defenders in the Digital Age" at 542.

[99] Amnesty International, Benetech, and The Engine Room, *DatNav*, p. 8.

practitioners introduce protracted review systems for social media communications and pay for the visibility of their social media messages, and thus the success of their digital communications depends on the resources of time and money. Risk perception can also silence, and unevenly so. Furthermore, as practitioners weigh up pluralism versus security in deciding whether or not to communicate digitally, erroneous risk perception can swing the balance too far to security.

What we have here, then, is a risk double bind – risk is bad for pluralism if you know about it, and it is bad for pluralism if you don't. If the latter, human rights practitioners are more likely to fall prey to communication risk. If the former, risk perception can prevent them from communicating digitally in the first place. This creates its own follow-on risk, like missing vital pieces of evidence or being dismissed as Luddites in the context of a broader pro-technology zeitgeist that has enthused donors. Though this double bind can make practitioners feel caught between paralysis and propulsion, it is not impervious to resistance. Next, I offer four approaches to loosening the silencing risk double bind.

VI LOOSENING THE SILENCING RISK DOUBLE BIND

The silencing risk double bind, constructed in part by commercial and political actors, threatens to squeeze the pluralism potential from human rights practitioners' adoption of ICTs. Political adversaries of the human rights world benefit directly from this silencing effect. The commercial actors of social media companies profit from human rights practitioners shaping their communications to social media logics, which can have silencing consequences. Human rights practitioners – as well as all those involved in the human rights and technology space, such as scholars, technologists, and donors – can, however, counteract these forces. In outlining four approaches to loosening the risk double bind, this chapter moves beyond the techno-pessimistic enumeration of materialized risk, and its potential contribution to silencing risk perception, toward a techno-pragmatic position. The four approaches, which work best in tandem, support the development and adoption of ICTs for human rights pluralism. The first pair of approaches, involving education and technology, are about mitigating materialized risks, while the second pair, involving reflexivity and discourse, relate to the construction and perception of risk. As human rights practitioners know very well, risk is an unavoidable element of their work; the aim is not to eliminate it, but to work alongside risk without it getting in the way.

A *The Educational Approach*

Knowing about risk without overblowing it involves understanding the origins of risk as well as mitigation strategies. Education projects for digital literacy – particularly around data literacy, security training, and social media advocacy – are proliferating

https://doi.org/10.1017/9781316838952.009 Published online by Cambridge University Press

apace with interest in digital human rights practices. For example, The Engine Room, Benetech, and Amnesty International recently published *DatNav: How to Navigate Digital Data for Human Rights Research*, which has since been translated into Spanish and Arabic.[100] Amnesty International's Citizen Evidence Lab walks practitioners through techniques for verifying digital truth-claims.[101] New Tactics in Human Rights hosts online conversations about using social media for advocacy, among other topics.[102] These educational resources are targeted at human rights practitioners, who share them through their networks of knowledge exchange.

That said, education has its limits. Expertise in digital fact-finding and advocacy can mitigate the materialization of some risk, but to the extent that the use of ICTs remains inscrutable – due, for example, to black-box algorithms or to an ever-shifting security terrain – some risk always remains. Furthermore, it is difficult to inform diffuse arrays of civilian witnesses about risk, which puts the burden of responsibility for digital security more squarely on the shoulders of human rights practitioners.[103]

B *The Technological Approach*

The technological pathway out of the risk double bind involves using ICTs built to address the risks engendered by digital communications. If human rights practitioners adopt these technologies to communicate with civilian witnesses, they go some way toward protecting those witnesses as well. For example, human rights practitioners are increasingly communicating via messaging applications, like WhatsApp, that are relatively impervious to surveillance. Many are consulting Security in-a-Box, developed by Front Line Defenders and the Tactical Technology Collective to introduce communities of users to digital security tools in seventeen languages.[104]

Of course, introducing technical fixes to digital communication risk may instead compound this risk, even if the technical fixes are done with the best of intentions. This is because the adoption of new technologies escalates the technological "arms race" between human rights practitioners and adversary state actors.[105] A case in point was the 2014 arrest for treason of human rights bloggers in Ethiopia, in which their use of Security in-a-Box was presented as evidence against them.[106] This potential for the inadvertent escalation of risks is one reason why the latter two

[100] Ibid.
[101] C. Koettl, "About & FAQ," Citizen Evidence Lab, 2014, www.citizenevidence.org/about/.
[102] "Using Social Networking for Innovative Advocacy," New Tactics in Human Rights, 2016, www.newtactics.org/conversation/using-social-networking-innovative-advocacy.
[103] McPherson, "Digital Human Rights Reporting by Civilian Witnesses," pp. 197–98.
[104] "Security-in-a-Box: Digital Security Tools and Tactics," https://securityinabox.org/en.
[105] Hankey and Ó Clunaigh, "Rethinking Risk and Security of Human Rights Defenders in the Digital Age" at 540.
[106] "Tactical Tech's and Front Line Defenders' statement on Zone 9 Bloggers," Tactical Tech, August 15, 2014, https://tacticaltech.org/news/tactical-techs-and-front-line-defenders-statement-zone-9-bloggers.

approaches, the reflexive and the discursive, are vital complements to the educational and technological approaches.

C *The Reflexive Approach*

Reflexive and discursive approaches call for critical perspectives on risk that unsettle taken-for-granted interpretations and practices. Reflexivity requires considering one's own role in making and perceiving risk, as well as the ways in which broader power relations shape risk assemblages.[107] It is all too easy to think about risk being an individual problem, when actually it is a socially constructed phenomenon.[108] For example, human rights practitioners are told to strengthen passwords, adopt encryption, and be vigilant about social engineering – or risk being hacked or surveilled. This is despite the fact that these risks emerge from the confluence of a multitude of commercial, criminal, and political actors.[109] Our tendency to individualize risk is to the benefit of these powerful actors. A broader view of risk that sheds light on these actors' roles redresses deniability and supports accountability in the determination of risk responsibility.[110] Furthermore, this view helps to safeguard individuals by painting a more comprehensive picture of risk and how and why it occurs.

Reflexivity about one's own roles and responsibilities in constructing risk is also important. Asking individuals to participate in a human rights technology project is also asking them to take on risk. This risk may be difficult to anticipate, in part because the context and resources where technologies are developed – usually the Global North – do not match the context in which the technology is being deployed. For example, digital security experts convinced one NGO to change its operating system, but the new operating system was not compatible with the NGO's printer. The NGO's employees had to bring files on memory sticks to printers at local Internet cafés. The memory sticks got lost in the process, which created a greater security risk than the original risk the operating system change was implemented to address.[111]

Practitioners in this sector must also be reflexive concerning their assumptions about civilian witnesses' participation in digital human rights fact-finding and these witnesses' knowledge of associated risk. Some civilian witnesses are driven to

[107] J. Kenway and J. McLeod, "Bourdieu's Reflexive Sociology and 'Spaces of Points of View': Whose Reflexivity, Which Perspective?" (2004) 25(4) *British Journal of Sociology of Education* 525–44 at 527.

[108] Deborah Lupton, *Risk*, 2nd ed. (London: Routledge, 2013) p. 21.

[109] Ulrich Beck, "The digital freedom risk: Too fragile an acknowledgment," openDemocracy, January 5, 2015, www.opendemocracy.net/can-europe-make-it/ulrich-beck/digital-freedom-risk-too-fragile-acknowledgment.

[110] Beck, *Risk Society*, p. 33.

[111] Z. Rahman, "Technology tools in human rights," The Engine Room, 2016, www.theengineroom.org/wp-content/uploads/2017/01/technology-tools-in-human-rights_high-quality.pdf.

document human rights violations by the somewhat idealized goal of speaking truth to power. For others, however, bearing witness may instead be a life-or-death matter, a matter of local or global politics, an exercise of identity, a function of resources – or simply a response to digital solicitations by human rights practitioners.[112] Some are accidental witnesses, while others are activists. Tailoring risk assessment to individual risk profiles and providing support for risk-bearing may require difficult, on-the-ground work that outweighs the mediation benefits of ICTs. Furthermore, practitioners may consider that soliciting digital information from civilian witnesses is too risky for certain contexts. Again, reflexivity is important, as practitioners need to consider whether they are or should be making silencing decisions on behalf of civilian witnesses. While an accidental witness may not have had an opportunity to think through risk, an activist witness's drive to digitally communicate documentation of violations may be underpinned by extremely sophisticated risk calculations.

D *The Discursive Approach*

The discursive pathway out of the risk double bind also involves focusing on the social construction of risk – this time by being aware of the possibility that actors communicate risks in order to control the behavior of others. In other words, risk perception can be a discourse used to protect or pursue power. The discursive approach to loosening the risk double bind involves identifying those who might benefit from risk discourses in order to assess how well perception corresponds to materialized risk.[113] For example, state actors may visibly surveil or punish digital activists not only to quell those individuals, but also to create a broader chilling effect on online human rights reporting.[114] As Amnesty International's secretary general stated following the UK government's 2015 admission that its agencies had been intercepting Amnesty's communications, "How can we be expected to carry out our crucial work around the world if human rights defenders and victims of abuses can now credibly believe their confidential correspondence with us is likely to end up in the hands of governments?"[115]

These discourses don't just serve political purposes; they can have commercial benefits, too. For example, tales of criminal and terrorist use of the "dark web" may

[112] M. Loveman, "High-Risk Collective Action: Defending Human Rights in Chile, Uruguay, and Argentina" (1998) 104(2) *American Journal of Sociology* 477–525; S. Madhok and S. M. Rai, "Agency, Injury, and Transgressive Politics in Neoliberal Times" (2012) 37(3) *Signs: Journal of Women in Culture and Society* 645–69 at 661.

[113] Lupton, "Introduction: Risk and Sociocultural Theory," pp. 4–5.

[114] K. E. Pearce and S. Kendzior, "Networked Authoritarianism and Social Media in Azerbaijan" (2012) 62(2) *Journal of Communication* 283–98.

[115] "UK surveillance Tribunal reveals the government spied on Amnesty International," Amnesty International, July 1, 2015, www.amnesty.org/en/latest/news/2015/07/uk-surveillance-tribunal-reveals-the-government-spied-on-amnesty-international/.

https://doi.org/10.1017/9781316838952.009 Published online by Cambridge University Press

arouse public suspicion about human rights practitioners' use of it in fact-finding, but they also sell newspapers.[116] Risk perceptions also create profit for the security sector, a major industry in which digital security is a growth niche.[117] The discursive approach to risk perceptions is particularly important, since, given the technical complexity of ICTs, most human rights practitioners must rely on external expertise to assess actual risk and appropriate responses.[118] Circling back to the education approach, incorporating this external knowledge must involve interrogating its motives.

VII CONCLUSION

Techno-optimism has surfaced in the human rights world, as in many others, based in part on the perceived benefits of ICTs for the pluralism of human rights communication. These benefits have been realized in a number of cases, but the application of ICTs has also materialized risk. As human rights practitioners consider whether and how to incorporate ICTs into their practices, this chapter has sought to outline some types of risk they may face and associated consequences for human rights pluralism. This risk, I argue, is a product of ICTs' affordance for mediation, or communication across time and place. This mediation, and the technical features it requires, alters the identity clues, interpretation cues, and contextual information communicators draw upon in order to increase the likelihood that their communication is successful.[119]

Furthermore, the use of ICTs introduces new intermediary actors to the human rights communication chain, and the technical complexity of ICTs makes these actors and their impact on communication more difficult to identify and assess.[120] Of particular note here are new commercial actors with profit motives. To be sure, human rights reporters have interacted with commercial motives before in their communication practices, such as in considering the marketability of newsworthiness decisions.[121] Never before, however, have commercial actors been so influential over and yet so hidden in mediation.[122] Cases in point are the commercial-political surveillance nexus, the lucrative gray market for spyware, and the proprietary, revenue-maximizing algorithms of social media platforms. Incorporating ICTs into human rights fact-finding and advocacy contributes to new risk assemblages for human rights practitioners.

[116] Beck, *Risk Society*, p. 46.
[117] Ibid., pp. 23, 46.
[118] Ibid., pp. 53–55.
[119] Thompson, *The Media and Modernity*, pp. 83–5.
[120] Beck, *Risk Society*, p. 22.
[121] McPherson, "Source Credibility as 'Information Subsidy'" at 333–35.
[122] Tufekci, "Algorithmic Harms Beyond Facebook and Google" at 208–09.

The types of risk outlined here are by no means the only ones that ICTs introduce or exacerbate for the human rights world. Others include the risk to human rights practitioners of secondary trauma brought on by exposure to images and videos of violations, or the retraumatization of individuals featured in advocacy material, particularly if the material is re-mediated and re-mixed.[123] The types of risk detailed here, however, have particular consequences for human rights pluralism. In digital fact-finding, human rights practitioners face surveillance risk that can imperil their physical security and deception risk that can jeopardize their reputational integrity. In digital advocacy, they encounter the risk of mistakes that have negative repercussions for reputations, as well as the risk that miscalculation poses for their resources. Some of these materialized risks and their repercussions silence human rights practitioners and civilian witnesses, while others deafen intended audiences to human rights communication. The perception of these risks can also be silencing, leading to a risk double bind in which both knowing and not knowing about risk can curtail human rights communication.

Acknowledging the silencing risk double bind throws into relief the importance of thinking about risk not in isolation, but rather as socially constructed. These social contexts produce values and connect individuals that could end up on opposite sides of a risk trade-off. In deciding whether or not to speak in the face of risk, human rights practitioners are choosing between the value of pluralism and the value of security. In so doing, they are also choosing between types of follow-on risk: the risk of physical harm and harm to reputations and resources if they choose pluralism, and the risk of ongoing human rights violations if they choose security. This means they are also making choices between risk populations.

The silencing risk double bind can feel unstoppable, part of the "juggernaut" of rapidly advancing technological change – with its associated complexities, inscrutable interconnections, and risk – that characterizes contemporary societies.[124] Yet, silencing is not inevitable. This chapter proposes four approaches to loosening the risk double bind: the educational and technological, which can limit materialized risk, and the reflexive and discursive, which can stay the construction of risk and erroneous risk perceptions. For practitioners, technologists, donors, and scholars, these approaches are useful heuristics for assessing risk. These heuristics also support human rights practices that allow successful digital communication to coexist with risk rather than be dictated by it.

The net impact of ICTs on the pluralism of the human rights world vis-à-vis the world of power it aims to hold to account is difficult to determine. What we can

[123] Bair, "Navigating the Ethics of Citizen Video" at 3; S. Dubberley, E. Griffin, and H. M. Bal, "Making Secondary Trauma a Primary Issue: A Study of Eyewitness Media and Vicarious Trauma on the Digital Frontline" Eyewitness Media Hub, 2015, http://eyewitnessmediahub.com/research/vicarious-trauma; Gregory, "Human Rights Made Visible."

[124] Beck, "The Digital Freedom Risk"; A. Giddens, *The Consequences of Modernity* (Cambridge: Polity Press, 1990), p. 139.

establish, however, is that materialized and perceived risk curtail pluralism unevenly within the human rights world. This dampening effect is stronger for human rights practitioners in more perilous political and social contexts and with less expertise and associated resources. It is not only particular organizations that are more affected by the materialized and perceived risk of digital human rights fact-finding and advocacy, but also the particular populations and particular human rights that they represent. For a world fundamentally concerned with pluralism, this momentum toward the use of technology creates risk for human rights enforcement in general, as it may be reinforcing inequalities around who speaks and gets heard on which human rights.