Technology in Society 82 (2025) 102779

Contents lists available at ScienceDirect

Technology in Society

journal homepage: www.elsevier.com/locate/techsoc

Digital recording and the hazards of unbounded moralized judgment

B.A. Kamphorst [a,*], E.R.H. O'Neill [b,1]

[a] Department of Media and Culture Studies, Utrecht University, Utrecht, the Netherlands
[b] Philosophy & Ethics, Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, the Netherlands

ARTICLE INFO

Keywords:
Digital recording
Moralized judgment
Norms
Privacy
Technology ethics

ABSTRACT

In today's techno-social environment, it is easy to make, store, and share digital recordings, such as photographs, audio fragments, and video streams, at an unprecedented scale. While there are often obvious immediate benefits to making and sharing digital recordings, serious hazards associated with these practices have thus far gone underappreciated. We contend that today's digital recording practices threaten to radically alter how we perceive and evaluate ourselves and others, producing an ongoing, socially and morally disruptive shift toward *unbounded moralized judgment*. The shift toward unbounded moralized judgment in turn poses several hazards, including widespread, difficult-to-restore reputation damage, negatively altered self-perceptions, and the stifling of morally right behavior. Our central claim is that in the current techno-social environment, every individual has a *pro tanto* reason to avoid being recorded and to avoid recording others. On the occasions where the reasons for recording outweigh those against, more must be done to counteract the hazards introduced by recording. We conclude the article by outlining possible avenues for technical, regulatory, and societal approaches to mitigating the hazards of unbounded moralized judgment.

> "The work offers the subjects the creation of an image of self, the distribution of which they cannot control, on a global scale and in perpetuity." (from C'mon C'mon; 56:25m)

1. Introduction

The ever-growing capacity for making, storing, and widely sharing high-fidelity digital recordings is changing the way in which we perceive and judge ourselves and others. Vivid audio and video replays of past actions keep recollections of past selves from degrading and open up the possibility for incurring praise and blame from an unprecedented number of people, over an unprecedented period of time, with an unprecedented degree of unpredictability. We argue that this phenomenon amounts to an ongoing, morally and socially disruptive[2] shift toward *unbounded* moralized judgment, brought about and sustained by socially disruptive recording technologies. The shift toward unbounded moralized judgment creates significant hazards. In particular, the expanding practices of making, storing, sharing, analyzing, searching, and using digital recordings in various contexts, risk inflicting difficult-to-restore damage to reputations (through moralized judgments by others) as well as negatively altered self-perceptions (through moralized judgments of oneself by oneself and via the influence of others' moralized judgments on one's own self-assessment). Moreover, we contend that for some people these two risks will be severe enough to stifle morally right behavior in contexts where such individuals are (or believe they are) recorded. We conclude that in the current technological ecosystem, every individual has a *pro tanto* reason to avoid being recorded and to avoid recording others. On the occasions where the reasons in favor of recording are substantial, more must be done to counteract the hazards of unbounded moralized judgment. With an eye to how the associated technologies are projected to advance in the future, we conclude by calling for the study and development of more creative technical, regulatory, and societal approaches to mitigating these issues.

The article is structured as follows. We begin by laying out some preliminaries in Section 2, explaining what we take the term "digital recordings" to denote and how these recordings differ from other types of representations, such as oral or written descriptions. We then proceed

* Corresponding author.
E-mail addresses: b.a.kamphorst@uu.nl (B.A. Kamphorst), e.r.h.oneill@tue.nl (E.R.H. O'Neill).
[1] The authors share first authorship as they have contributed equally to this work.
[2] On the concept of social disruption, see [80,81].

https://doi.org/10.1016/j.techsoc.2024.102779
Received 10 June 2024; Received in revised form 29 September 2024; Accepted 1 December 2024
Available online 2 December 2024

to describe the digital ecosystem in which digital recordings are embedded and its associated properties. Lastly, we introduce our understanding of "unboundedness" in relation to moralized judgments.

In Section 3, we discuss how the properties of the current digital recording ecosystem are facilitating unprecedented levels of making, sharing, and long-term storing of recordings. We claim that these practices are leading to significant change in the form and scale of moralized judgment, via four key, interlocking trends: trends toward repeated exposure, scaled exposure, unpredictable exposure, and unpredictable modification and use. These trends together, we argue, are facilitating a shift toward unbounded moralized judgment, which in turn creates three related but distinct hazards.

In Section 4 we expound on what these hazards entail, building on pertinent past work from privacy scholars. The first hazard is that the temporally open-ended and widespread availability of digital recordings can lead to permanent stigmatization by (many) others. The second hazard is that frequent exposure to one's own past recordings, together with others' reactions to those recordings, can adversely affect how one views and evaluates oneself. Here, the focal point is how one's self-perceptions are shaped in response to recurring or continued moralized judgments, including one's own, about past actions. The third and final hazard we consider relates to one of the ways in which people may respond to the risk of reputation damage or negatively altered self-image. We argue that the risk of backlash can lead to morally problematic behavior in contexts where people believe recordings are being made. In Section 5 we look forward and anticipate how some further technological developments may aggravate the risks associated with unbounded moralized judgment.

Finally, having laid out our concerns, we turn to possible mitigation strategies in Section 6. There, we sketch potential societal, regulatory, and technical mitigation strategies, and call for sustained reflection on these approaches as well as the hazards we have identified. Section 7 concludes.

2. Preliminaries

In this section we set the stage for our main argument by outlining the conceptual boundaries of the relevant terms. We begin by detailing our understanding of the term "digital recording."

2.1. What are digital recordings?

We use the term "digital recording" to denote persisting digital data that is produced via a process in which analog audio or visual information from events is captured and digitally encoded, affording the possibility of creating auditory or visual representations of those events.[3] As would be the case with analog recordings (e.g., audio or visual information that is captured on celluloid-based film or magnetic tape), we take the resulting representations to have a likeness[4] to the events, but a likeness that may, up to some threshold,[5] be imperfect—be it because of inadvertent loss in the recording process or because of global settings of the recording device that determine which information is collected (e.g., shutter speed settings that determine how much light reaches a camera sensor). That said, we assume that many of today's high-fidelity digital recordings result in representations with a high degree of verisimilitude.[6] Our focus in this article is on the sub-class of digital recordings that depict identifiable individuals or groups.[7]

We focus on digital recordings but acknowledge that some of the worries we will present in this article also pertain, at least to a certain extent, to digital preservations of written statements by an individual or to descriptions of an individual. Textual or testimonial evidence can certainly form the basis for normative judgment, and the long-term preservation and widespread availability of such evidence can potentially lead to similar issues as the ones we will present here. However, we contend that there are several important differences between likenesses on the one hand, and written statements or descriptions on the other, which make us more concerned with the former. First, we submit that humans are prone to make a broader set of judgments (for instance, judgments about the subject's bodily features, tone of voice, attitude, character, intentions, actions, mannerisms, etc.)[8] on the basis of a likeness (above all, a high-definition, photorealistic video, and to a lesser extent a still photo or an audio recording) than on the basis of description or written statement. Second, in comparison with a description, there is a degree of apparent verisimilitude associated with mechanically produced likenesses, especially when they are high definition.[9] One perceives the agent's (reproduced) properties and actions with one's own eyes or ears.[10] Third, the high degree of likeness that a high-definition audio fragment, photo, or video has to an individual or group facilitates (re)identification of that individual or group in a way that

[3] While our definition strictly excludes analog recordings, it does count a digitized version of an analog recording as a digital recording. In such cases, analog audio and visual information from events is initially captured through a non-digital recording process and only subsequently digitally encoded. In the current socio-technical context, this type of recording object can pose many of the same problems as recordings that are produced more directly—they can be just as easily transmitted, shared, duplicated, stored, modified, searched, and analyzed.

[4] Our use of the term "likeness" here should not be controversial; though we recognize the arguments against relying on the notion of likeness or resemblance for constituting a general theory of representation (e.g., [164]), we are merely subscribing to the established idea that likeness or resemblance can help explain what is distinct about pictorial representations (as opposed to textual representations). In the present context, these notions are thus helpful for distinguishing the kinds of representations we are focusing on (pictures, videos, but also audio) from various textual representations (e.g., blogs, tweets, posts, etc.).

[5] We introduce this threshold to indicate that while the resulting representations need not have a perfect likeness to the events they are representations of, we presume the representations to reflect, in some meaningful sense, the events as they unfolded. How this threshold should be determined, though, is a question about which we remain agnostic. For our present purposes, we simply assume that at some degree of distortion in what is captured or in how it is encoded, the resulting data no longer affords the creation of a representation of the original event.

[6] Of course, the verisimilitude can be altered post hoc by manipulating the data, for example by modifying the pitch of the audio or tweaking the hue, saturation, or lightness of an image or video during playback. Likewise, data may be augmented or enriched, for example through the use of filters that transform (aspects of) the audio or video. We return to these types of modifications below.

[7] What kinds of depictions make a person identifiable change over time. For instance, machine learning may make it possible in the future to identify individuals on the basis of a recording of how they walk or an image of their masked face, etc.

[8] Video and audio in particular can capture mannerisms in a way that static imagery cannot. On the difference between recordings and descriptions, see also Tunick [27], who writes, "A recording usually reveals details that can't readily or as effectively be conveyed merely by reporting" and "… the recording allows a different kind of exposure. It doesn't merely convey information; it gives the audience an experience of its subject" (p. 658).

[9] This point is one with lengthy pedigree in film theory (e.g., [82]). See the discussion in Atencia-Linares & Artiga [83] on the philosophical view that mechanically produced images are epistemically privileged.

[10] Expanding on an argument from Dan Cavedon-Taylor about photographs, Rini [3] argues that video and photographs can offer perceptual evidence. The contrast these authors are drawing is with hand-made media, which they say can offer only something like testimonial evidence. The apparent verisimilitude factor of digital recordings may be undermined in the future if deepfake media continues to proliferate (see e.g., Fallis [153] on how the reliability of photographs and video may be undermined, as well as Rini [3]), but we take it that it has not been undermined yet. Until it is, the apparent verisimilitude factor will mean that likenesses pose a bigger threat than written statements or descriptions, with regard to unbounded moralized judgment.

description does not. Consequently, it is audio fragments, images, and videos that bear a likeness to particular individuals or groups that we are most concerned with in this article.[11]

Lastly, a comment on the relation between digital recordings and "deepfakes." Digital recordings can serve as input for the generation of partially synthetic digital media such as internet memes [1] and "deepfake" videos [2]. We take it that synthetic media of those kinds are themselves not recordings—(the degree to which they involve fabrication or modification of recordings exceeds some relevant threshold)—but we view it as an important feature of digital recordings that they play such an important role in generating such media. Depending on the source material used, partially synthetic digital media can *appear as if* they are recordings—portraying individuals as doing or saying something they did not in fact say or do or as possessing properties that they do not have (cf. [3]), and thereby providing subjects for judgments. As Millière puts it, a "convincing face-swapping deepfake may be indistinguishable from a genuine [recorded] video" ([2], pp. 19–20). As will become clear, these possible fabrications from or modifications of actual recordings contribute to the hazards we identify.

2.2. The digital ecosystem

A second set of preliminary remarks pertain to the context within which digital recordings result in hazards. We wish to underline that it is not the digital recordings themselves that, in isolation, produce the shift toward unbounded moralized judgments and the hazards that follow. Rather, it is their interplay with the "digital ecosystem" (as well as the social environment) in which they are embedded that does so. And while it would be an unproductive exercise (not to mention a Sisyphean task) to exhaustively list all low-level technologies comprising this ecosystem, it will be instructive to look at a number of its key, high-level components in order to bring into view the properties that contribute to the hazards of unbounded moralized judgments.

Viewed from a sufficiently high level of abstraction, and in relation specifically to the hazards we are concerned with, we can identify five key components of the digital recording ecosystem, related to recording, storage, duplication and sharing, modification, and search and analysis. First, as we have seen, there are the digital recording technologies themselves that allow for events to be captured and replayed. These include all camera-fitted and microphone-fitted devices that capture audio and/or visual information and preserve it as "archival media" on digital storage media such as optical disks, magnetic disks, or flash drives. Examples include smartphones, CCTV cameras, home security systems and camera-fitted "smart" doorbells, compact action cameras (like GoPro), and spycams.[12] In this category we also include the (software) technologies that make it possible to make screen recordings (e.g., of video conferencing calls) and to surreptitiously activate cellphone microphones or cameras (e.g., Pegasus).

Second, there is the class of technologies that allows audio fragments, images, and videos to be stored for long periods of time, in a multitude of formats, and in various locations across the globe. Included in this class are the hardware components that enable the long-term storage of media files and their distribution over the internet (e.g., hard drives, flash drives, but also (file) servers, routers, switches, fiber optic cables, etc.), as well as (low-level) software components that determine how data is represented on disk and how data can be exchanged (operating systems, file transfer protocols, compression algorithms, etc.). Major commercial cloud storage service providers have built their businesses on top of these technologies and offer vast amounts of storage space, often with (semi-)automatic backup and replication mechanisms in place, at low cost for the end user. In addition, local file storage products such as USB flash drives, internal or external hard drives, and even complete network-attached storage (NAS) systems have become mainstream and also offer significant "offline" storage space at affordable prices (especially in middle- to high-income countries). This abundance of easily accessible storage space has lowered the threshold for holding on to digital recordings for extended periods of time (potentially spanning generations). Moreover, the interplay between online and offline storage for digital recordings is pernicious when it comes to ensuring the deletion of recordings, for there is often no way of knowing whether a copy of a recording is still stored in a company's backup system (potentially residing in a different country where different laws apply) or on an individual's workstation or flash drive.

Third, there is the class of web-based services, including cloud storage products and file sharing services, that form the infrastructure for media files to be quickly and easily duplicated and shared at little to no financial cost to the end user. Included in this class are social media platforms as well as digital marketplaces where digital content can be shared, traded, or sold. This infrastructure and these services contribute to an ecosystem in which recordings can be distributed globally at high speed and stored in locations unknown to either the creator of the content or the subject portrayed. Moreover, the business models of some of these platforms, in which the creation of new content and the widespread sharing of that content is incentivized through monetary and social rewards, contribute to the growing practice—and, to a certain extent, the normalization—of recording one's own or other people's actions and putting those recordings online for (many, potentially unknown) others to see.[13]

Fourth, there is a class of technologies that facilitate the modifiability or malleability of digital recordings. Due to the nature of how audio and visual information is represented in various digital formats, media files can typically be altered—and with little skill and cost, compared to many other media formats.[14] Today, even the most basic smartphones

[11] One might furthermore wonder whether our arguments about digital recordings also apply to non-digital representations of humans, such as lifelike drawings, paintings, statues, mimicry, and so on. One reason we are more concerned with digital recordings than these other representations is that in the modern digital recording ecosystem, the former are much easier than the latter to create, store, duplicate, share, modify, search, and analyze; another reason has to do with the lesser degree of apparent verisimilitude typically possessed by the non-digital representations.

[12] In a way, these technologies can be thought of as successors to the analog "snap camera," the spread of which was a source of worries for legal scholars Samuel Warren and Louis Brandeis in 1890. Their seminal article "The Right to Privacy" [84] sounded an early warning about the hazards of recording. Today's digital counterparts have made it easier than ever to capture audio information, images, and (extended) videos at any time or place, overtly or covertly.

[13] Though most large platforms do implement various strategies to counter potential misuse, e.g., fine-grained permission schemes, download protection algorithms, or digital rights management (DRM) technology, most of these protection strategies can technically be circumvented (cf. [152]). Most platforms also offer "report abuse" mechanisms, but by the time an abuse report has been filed, the recording in question is typically already downloaded and stored offline. Thus, even if the platform takes timely action and removes the reported content from their platform, that content may (straightaway or in due time) resurface somewhere else.

[14] A number of scholars have observed that in the past, too, it has been possible to modify representations of subjects: with paintings and other visual art, some individuals have the skill to alter artwork while retaining a sufficient likeness of the individual depicted; with analog photography and videography there also exist techniques to distort the recording as it is made, and to alter the recording after it has been made [85]. In response to this, we stress that there is a vast difference between digital recording technologies and past representational technologies with regard to the level of expertise required to make modifications, the ease, cost, and speed with which any given individual can make modifications, and the scale at which modified representations can be copied and distributed. There are also important differences in the types of modifications that one can easily make—e.g., it has not previously been feasible for individuals to make realistic pornographic films featuring (unconsenting) people they know ([5,86]; see also [87]).

have the capability to modify images and videos, e.g., by cropping or by changing the hue, or by creating overlays containing custom drawings or text. Specialized software suites allow for further modification and customization by giving end users free range to change (aspects of) the environment in which a subject of an image or video was originally positioned. This allows for radical transformations to occur where subjects of a particular recording may be placed in entirely new contexts. The same is true for modification of the recorded subject: augmented reality filters for making a subject appear older or younger, adding makeup, "beautification," etc. are widely available on popular social media applications [4]; likewise, AI-based software makes it easy to modify facial expressions of recorded subjects. Generative AI, combined with social networks through which individuals learn how to wield such technologies, have made it easy for people with limited technical skills to generate synthetic digital media, including pornographic material, of ordinary people they know, on the basis of one or a few real recordings of that targeted individual [5–7].

Fifth, advancements in search and analysis capacities, largely due to improvements of scaling techniques and machine learning algorithms (e.g., used for face and voice recognition), have introduced an unprecedented capacity for ordinary people to selectively locate recordings of particular individuals on the basis of relatively little information. Several companies have already offered services by which one can upload a few photos of an individual and retrieve a vast trove of other images of the individual from across the internet ([8–10]; see also discussion [11,12]). In a previous decade, when putting a photo or video on an isolated website, it might have been reasonable to count on it remaining undiscovered in virtue of its being a needle in a haystack—this is no longer the case.[15]

The digital ecosystem as described above consists of numerous technological innovations in hardware and software that are contingent: many, if not all, aspects of the ecosystem could have been implemented differently, and quite possibly will be implemented differently in the future. Moreover, it is undoubtedly possible to refine the five categories we have used to sketch the ecosystem. Still, our description of the ecosystem should be sufficient to bring across the general idea that the hazards we identify with digital recordings are tied to the socio-technical environment in which the recordings are embedded. The hazards result not just from the fact that recordings can be made, but that they can be easily stored for long periods of time, duplicated, shared, searched, and used in a wide array of (unexpected) manners, all at unprecedented speed and scale.[16]

Stating that these features of the ecosystem contribute to the hazards associated with digital recordings does not mean they do not also have immediate and long-term benefits. The value to storing and sharing digital audio and media files—whether helping to preserve memories, aid in education, or keep faraway friends connected—is self-evident. Likewise, the possibility to modify media files has a host of possible uses, some benevolent (e.g., adding a missing family member into a family photo, in a context where historical accuracy is not needed), some satirical or entertaining (e.g., poking fun at a world leader), some artistic. The practice of altering media files through software as a creative enterprise can also positively contribute to community building or political engagement. It is important to note, therefore, that we do not wish to cast a negative light on the practices of creating, duplicating, storing, sharing, searching, or altering recordings as such. Rather, what we wish to draw attention to is the way in which digital recordings, embedded in the current digital ecosystem, facilitate a shift toward what we call "unbounded moralized judgment." As we will argue, it is this shift that raises the hazards that we are concerned with. Before getting to the argument itself, however, let us conclude this section by explicating how we use the term "moralized judgment" in the present context and how we understand the notion of "unboundedness."

2.3. Moralized judgments

We claim in this article that the practices surrounding modern digital recording are disrupting our ordinary practices of normative judgment—practices of judgment formation, retention, sharing, revision, etc. The category of normative judgments incorporates a great variety of judgments, including judgments that something is right, wrong, permissible, obligatory, good, bad, praiseworthy, admirable, evil, improper, hideous, disgusting, and so on. Some normative judgments are more interesting than others, though, in the sense that they call for action in more contexts, or allow for fewer exceptions, or produce more motivational force in the bearer. The class of normative judgments that most concern us are those that have one or more aspects traditionally associated with moral judgments and that are such that the person making the normative judgment is likely to take action in light of their judgment. The label we will use to refer to this class of judgments is "moralized judgments."[17] Many different aspects of judgments can (indirectly and in different ways) influence whether a person will take action. We take it as plausible that, among other things, a person will be more likely to take action in light of their judgment to the extent that the judger categorizes the judgment as a moral judgment (as opposed to some other type of normative judgment),[18] the judger views actions in violation of the judgment as serious violations,[19] the judger believes

[15] Cf. Hartzog [88], pp. 1038–1042) and Hartzog & Selinger [89] on obscurity and the often-underappreciated benefits that it has provided in the past; see also Allen [90] on the disappearance of the anonymity of the crowd.

[16] What we are saying about digital recording mirrors points that others have made previously in the privacy literature, regarding how changes in information technology, digitalization, the arrival of the internet, and so on, have led to an increase in the ease of collecting, storing, sharing, and analyzing data, which have significant consequences for whether traditional ways of doing things still suffice to protect privacy (see e.g., [91]).

[17] We do not use the term "moral judgments," because whether it is possible to carve off "moral judgments" as a distinct kind of normative judgments (i.e., as distinct from conventional, aesthetic, epistemic, or other varieties of normative judgment) is highly contested [13, 92–97, 160]. There are numerous, conflicting proposals for what, if anything, distinguishes moral from other types of normative judgments and norms. We believe it is not necessary to weigh in on the question of whether moral judgments are a natural kind in order to convey our point that for one influential set of normative judgments, i.e., the set which we label "moralized judgments," the modern digital recording socio-technical ecosystem is producing significant changes to human practices pertaining to those judgments. Admittedly, there is also some dispute about what it is for a judgment to be "moralized." Brady et al. [98] offer a content-based account of moralization of content: "we classify content as moralized if it references ideas, objects, or events typically construed in terms of the interests or good of a unit larger than the individual (e.g., society, culture, one's social network)" (p. 978); Rozin [99] offered a broader account of moralization of activities, entities, etc.: "Moralization is the process through which preferences are converted into values, both in individual lives and at the level of culture" (p. 218). Our account is different than either of these. However, we take it that the term "moralized judgment" has fewer connotations and makes a better candidate for repurposing as a term of art in the way that we propose.

[18] Wright [100] suggests that humans "use the classification of 'moral' to mark those beliefs, values, and practices that they view as unacceptable forms of deviance" (p. 88); she summarizes past research as finding that people have "the most attitudinal and behavioral intolerance for divergence that they have classified as moral" (p. 87) and that people "were most willing to prohibit, censor, shun, and punish divergent moral beliefs, values, and practices" (in comparison with divergent personal or social beliefs, etc.) (p. 88). See also [101–103].

[19] The degree to which the person takes violations seriously may be indicated by the degree to which the person tends to view violations as worthy of sanction, or the person is disposed to take action to discourage such violations, e.g., via expressed disapproval, gossip, avoidance of the perpetrator, or more elaborate sanctions.

their judgment also applies to similar cases at a wide range of other times and places,[20] the judger (implicitly) takes the judgment's truth or falsity to be determined by something other than their own mental states or the decree of an authority,[21] the judgment is accompanied by any of the emotions that are canonically characterized as moral emotions (e.g., righteous anger or indignation, outrage, moral disgust, contempt, admiration, shame, guilt, pride, etc.),[22] and the judger takes their belief on the question to be an important part of who they are, such that if they had a different view, they would be less authentically themselves or would be in some sense a different person.[23] The more of these features a normative judgment has, and, in the case of features that come in degrees, the greater the degree to which the judgment has a given feature, the more "moralized" we will say that the judgment is.[24]

Importantly, for a normative judgment to be moralized in our sense, it need not involve any of the topics that many psychologists in the twentieth century associated with morality, such as harm, justice, or rights [13]. That is, the account of moralization that we are using is content neutral. This allows us to communicate a point about normative diversity that is crucial for our argument: even if there is a set of norms that is frequently moralized across cultures (i.e., matters relating to harm and care, distributions of resources and punishments, etc.),[25] beyond that domain of overlap there is also a vast diversity in the substance of other norms and normative judgments that cultures, sub-cultures, and individuals moralize [14].[26] Such diversity is an important contributor to the hazards we will discuss. As Fiske & Rai [15] have argued, many people who commit heinous violence are motivated by what they believe to be moral reasons. Likewise, in cases of networked harassment online, harassers across the ideological spectrum often justify their actions by appealing to (what they appear to view as) moral reasons [16]. It is thus important to keep in mind that moralized judgments can include not only judgments like "that was racist," or "that was selfish," but also potentially judgments like "that person deserves what's coming to them," "that person doesn't know their place—they ought to be knocked down a few notches," and "people like that should be wiped off the face of the earth."

It should be clear by this point that by "moralized judgments" we do not mean "moral judgments that are right." At the same time, our claim is not that all moralized judgments are bad. Moralized judgments have played a multitude of critical roles in supporting social life and helping individuals and groups achieve goals.[27] Our contention in this article is that in our substantially altered techno-social environment, moralized judgment practices are changing in such a way that they may no longer serve many of the beneficial functions that they sometimes served in the past and instead that they are producing a suite of hazards that are as of yet underappreciated. Historically, moralized judgment practices have been substantially constrained and shaped by human cognitive faculties and inherited cultural traditions. We claim that the technological changes discussed in the previous section, in particular the advancement in storage, sharing, search and analysis capacities, and the developments in synthetic digital media (e.g., deepfakes), are profoundly altering our practices of moralized judgment by exposing people—through their digital representations—to judgments by an unknown many others, at unexpected moments and in unforeseen ways, for potentially indefinite periods of time. This is what we consider the shift toward *unbounded* moralized judgment.

2.4. The notion of unboundedness

By "unboundedness," we refer to a lack of principled or natural boundary conditions for when something stops being open to judgment (by others or by oneself). In brief, the term is intended to signify a state in which actions, individuals, or states of affairs can be judged continuously and indefinitely by an indefinite set of people. Presumably, in a strict sense, given physical constraints, a state of complete unboundedness cannot obtain; consequently, we restrict ourselves to speaking about *a shift toward* unboundedness. It is the (radical) weakening of boundaries that is leading to the hazards we identify.

There are two further observations to be made about this. The first is that talk about a shift *toward* unboundedness indicates a continuum, in the sense that any assessment of how close we are to unboundedness of moralized judgment is a matter of degree. This continuum allows us to acknowledge how our practices around moralized judgments have changed from how they were in the past, and to make statements about how, depending on the choices we as a society make in regards the technological ecosystem and our socio-cultural environment, things may evolve for better or worse (in this regard, see also Section 5).

The second observation is that we consider the shift toward unboundedness in different dimensions. The primary dimension is time (having recorded actions be subject to judgment for long periods of time), but we also consider the number of potential judgers as a relevant dimension, as well as the myriad ways in which content, original or modified, may be subjected to scrutiny. This means that scenarios may vary in degree across dimensions as well, allowing for more fine-grained analyses of what certain interventions to prevent the hazards of unbounded moralized judgments may accomplish (e.g., implementing technical limitations on sharing may not change how long files may be

[20] In discussions on what distinguishes moral judgments and norms from other types of normative judgments and norms, this property is usually characterized as generality or universality (e.g., in Turiel [104]).

[21] Cf. discussion on this topic in Wright [100], O'Neill and Machery [14]. See also Stanford [105] on the externalization of norms.

[22] See e.g., Haidt [106] for one overview on moral emotions.

[23] See e.g., Strohminger & Nichols [107] on the relationship between values and identity. Another factor that may make a difference is whether the judger (likely implicitly) takes the judgment to be such that two people cannot disagree about the judgment without one being wrong, and where what is at issue is not merely a matter of taste. Goodwin & Darley [108] discuss this property under the label of "objectivity." For discussion of several of these features of normative judgments, see e.g., Levine et al. [109], O'Neill [110], Kumar [111]. We adapt the notions of seriousness, generality, and mind- and authority-independence from the literature on the moral-conventional distinction [104], according to which (on some versions of the distinction) moral norms are viewed as serious, taken to apply universally, and are authority independent, in contrast to conventional norms.

[24] Conceivably, one could operationalize the dimensions of this concept of moralized judgment and construct an overall measure of moralization, allowing one to compare the degree to which a particular individual moralizes different judgments and to compare the degree to which different individuals moralize a given judgment. If a plausible formal model of the concept could be achieved, then perhaps it could be combined with other formal models of human normative cognition and decision-making, such as Kleiman-Weiner et al.'s model of moral learning [165] or Capraro & Perc's model of personal norms [112], to facilitate further empirical research on how normative judgments with various features influence action.

[25] Two of the most prominent psychological theories of human morality, i.e., Moral Foundations Theory [113,114] and Morality as Cooperation Theory [115,116], each propose that there are some values—which they characterize as moral values—that appear in all cultures, even if they are given differing levels of importance in different cultures.

[26] It is likely that within the set of judgments that we would characterize as highly moralized, there are some judgments that others would classify not as moral but instead as aesthetic, epistemic, religious or some other category. Inasmuch as those judgments fit our criteria and thus are likely to influence action, they are of interest to us, regardless of how others categorize them. We readily acknowledge that changes in the digital ecosystem may also be altering practices involving judgments with *few* of the dimensions we use to define moralization; changes in practices related to these other types of normative judgments may also warrant study, but we leave that as a question for future research.

[27] For arguments about the various roles that human normativity and morality may have played in human evolutionary history, see e.g., Sterelny [117], Pettit [118], Tomasello [119].

stored for, and vice versa).

These final observations conclude our preliminary remarks. Let us turn to the developments that are driving this shift toward unbounded moralized judgment.

3. Shifting toward unbounded moralized judgment: Four interlocking trends

In the preliminaries, we provided a sketch of the digital ecosystem in which digital recordings are embedded, and identified five key features of this ecosystem—changes in recording, storage, duplication and sharing, modification, and search and analysis technologies. In this section, we will argue that there are four interlocking trends supervening on these features that together are driving the shift toward unbounded moralized judgment. These trends concern the way in which digital recordings can be 1) repeatedly retrieved and resubmitted to critical scrutiny and moralized judgment at an unprecedented scale by 2) an unprecedented number of individuals (on a global scale and over time) at 3) unprecedentedly unknown and unexpected times in 4) unprecedentedly unforeseen and unexpected forms. Collectively, they constitute a substantial set of changes to human practices of moralized judgment. We will discuss each of these forces in turn.

The Repeated Exposure Thesis. Inherent to the preservation of audio and video information in recordings is the possibility for reproduction and playback. This possibility was already present with mechanical analog recordings (e.g., reproducing sounds on the basis of inscriptions made by early phonographs that captured air pressure changes) but with modern-day's digital recordings, playback is easier and of a higher resolution than ever before. Moreover, the ways in which digital recordings are typically stored and duplicated in today's digital ecosystem means there is much less risk of recordings naturally degrading with time. This robustness, together with the ease and low cost of making and storing recordings, and supported by large data banks and initiatives such as the Internet Archive Way Back Machine that aim to preserve snapshots of digital content for posterity, make it so recorded actions can potentially be retrieved and (re-)submitted to judgment, time and again.[28] The result is a considerable contrast with historical contexts in which human practices of normative judgment about past events were based solely on memory, oral tradition, and limited imagery. This constitutes the first of the ways in which moralized judgment practices are undergoing radical and disruptive change: as people's actions are increasingly being recorded, more actions thereby remain open to—or can be resubmitted to—moralized judgment (by the recorded individuals themselves or others) long after the actions were performed.

The Scaled Exposure Thesis. When digital recordings are resubmitted to moralized judgment, the exposure is typically not restricted to the original group of participants, onlookers, and witnesses present at the original event. Rather, what the digital ecosystem allows for—and, in many cases, incentivizes—is the rapid, widespread sharing of digital recordings to large, new audiences. Depending on circumstances, digital content that goes "viral" can reach millions of viewers, including friends, family members, neighbors, co-workers, stalkers, etc., in a matter of minutes (cf. [17,18]). This way of broadcasting information about one's own and other people's actions is changing moralized judgment practices by providing a potentially very large set of individuals, who otherwise would not have known about the event, with the opportunity to form judgments about the subject or subjects portrayed in the recording.[29] Recorded subjects in these situations are thus exposed to judgment by substantially more people than they typically would have been historically, often including judgment by many people who lack relevant contextual information. After all, any encounter with a digital recording in effect involves the consideration of an event outside the context of the original, recorded event.[30] Relatedly, the phenomenon of "context collapse" [19] on the internet produces conditions in which multiple audiences and people playing many different roles encounter the same digital recording. Moreover, the subjects in these recordings are likely to be judged by many individuals who they themselves do not know, creating a potentially pernicious asymmetry where the subjects themselves are unaware of who has had access to the digital recording and where their possibilities for anticipating negative consequences or obtaining redress for harms done are therefore limited. Here the diversity of what humans moralize—around the world and over time—matters a great deal. A person could be the most admirable of moral agents and still attract moralized criticism from individuals and groups around the world.

The Unpredictable Exposure Thesis. We have already mentioned that, as a result of the speed at which digital recordings can spread, moralized judgment can occur and become widespread unusually quickly. This, however, is only one end of the unpredictable exposure time spectrum; on the other end is exposure that occurs many years in the future, long after the subject has forgotten about the event, or even years after their death. In between are a multitude of other moments in which a recording may unexpectedly resurface or attract new attention.[31] As such, the subjects of these recordings can be subjected to new moralized judgments by many (unknown) others at times when they least expect it. In this regard, the digital ecosystem in which digital recordings are embedded currently facilitates a climate of insecurity about the consequences of one's actions, and their normative status in the eyes of others, that did not exist to the same extent before.

The Unpredictability of Modification and Use Thesis. The future uses of modern digital recordings are difficult if not impossible to accurately anticipate. One reason for this is the unpredictability of how digital content may be modified *post hoc*: today's digital editing tools, increasingly aided by AI techniques, enable radical transformations of recorded contents. As we have already discussed, digital recordings may be employed to create deepfake photos, videos, and audio; filters and other editing techniques may alter portions of what is depicted in the original recording. In addition, snippets and stills can easily be removed from the fuller context of the source material in which they originated. Any video recording, whether made by a commercial production company or an ordinary person, is now a storehouse of clips and stills that can be extracted with little effort, to be repurposed for static memes and gifs.

Closely tied in with this is the unprecedented unpredictability of the *purposes* for which digital contents may be employed. Some unpredictability in use, in itself, is not new: voluntary recordings, for example of one's sexual activities, could be used, in their unmodified state, not only for the participants' own entertainment but also for purposes of extortion or blackmail. Scaled exposure, though, dramatically increases the set of people capable of doing things with a given recording, thereby dramatically increasing the set of purposes for which a given recording

[28] See related discussion in Frye [120] regarding the "virtually permanent status" of shaming on social media (p. 142).

[29] Though the process of broadcasting information is not new—TV, radio, and print media already had the power to amplify certain happenings—the scale at which such broadcasting is happening is unprecedented, in terms of audience, the parties broadcasting, and the range of events being broadcast.

[30] We thank Dawa Ometto for emphasizing this point in conversation. Patton [121], commenting on recording in the context of surveillance, makes an important observation on this topic: "The introduction of electronic surveillance to a place makes the events of that place accessible to any number of other places and times. Participants in the social context thereby have their ability to read the circumstances of that place's context diminished. Surveillance introduces an ambiguity into the place whereby people become less clear on who their actions are accessible to and in what circumstances their actions may be reviewed." (p. 184).

[31] Many subjects of viral attention and online shaming describe how attention comes in waves, dying down for a while before rising again.

might be reused. Furthermore, the ever-growing possibilities for modification greatly expand the set of purposes a recording can be used for, inasmuch as aspects of the recording that link it to its original context can be removed entirely and new elements (objects, surroundings, people, etc.) can be injected. This trend is disruptive to moralized judgment practices in the sense that people are now much more easily subjected to moralized judgments for properties they did not possess or actions they never performed, as well as subjected to moralized judgments held with a high degree of confidence (due to being based on people's own perceptions of the content) yet hinging on misinterpretations of what occurred.

We contend that, taken together, these four forces are facilitating a socio-technical climate in which recording human persons raises substantial hazards. The shift toward unboundedness constitutes a dramatic disruption to the conditions under which humans have made moralized judgments in the past. Yet thus far the significance of this has gone underappreciated both by academics and by the public. In what follows, we shall expound three hazards of unbounded moralized judgment, which we think deserve more serious consideration in societal reflections on how the digital society should be given shape.

4. Three hazards

The picture emerging from the previous section suggests that in the present socio-technical climate, recording increasingly exposes the subject to unbounded moralized judgment. This raises serious worries about broad, difficult-to-restore reputation damage, negatively altered self-perceptions, and even the stifling of morally right behavior. We shall discuss these hazards in order.

4.1. Tarnished reputations

Reputation, Solove reminds us, "is a currency through which we interact with each other" ([20], p. 160). Seen as a public representation of other people's opinions about one's credibility with respect to certain traits, reputation plays a key role in establishing and maintaining interpersonal trust ([21]; see also [22]). As such, reputation co-determines one's social standing, from the perspective of various groups, and, with it, how one is treated in society (cf. [23,24]). Even when one places little stock in others' opinions of oneself, reputational harms can have a broad range of practical ramifications, in the sense that they may "impair a person's ability to maintain 'personal esteem in the eyes of others' and can taint a person's image in the community" ([20], in reference to [25]). Having one's public image tarnished in this way is liable, in turn, to influence the opportunities one receives and the obstacles one faces. Once perceived as having acted wrongly or as having taken the wrong position on a moralized issue, individuals may be excluded from consideration for employment, friendships, relationships, housing, and other important social interactions. At the extreme—as has been amply demonstrated over the past decade—people may encounter doxing, swatting, harassing phone calls or home visits, unwanted confrontation with material intended to shock or disturb, death threats, orchestrated efforts to make them unhireable, efforts to have their children removed from their home, and nonconsensual and debasing use of personal imagery, among other potential consequences.[32]

Due to their socially constructed, dynamic nature, reputations are known to be fragile [21]. Examples abound of people—celebrities and ordinary individuals alike—whose good reputations were substantially undercut once certain information became public. We do not claim that the social mechanism underlying reputational penalizing is problematic in and of itself—one might even think it is inherent to sociality. Given that reputations inform decisions about whom to trust, presumably it is a desirable feature of reputation that it should fluctuate in response to one's actions. In that sense, reputational harm is sometimes simply warranted. Digital recordings, however, together with the digital ecosystem in which they are embedded, contribute to a problematic increase of reputational fragility in (at least) five related but distinct ways.[33]

First, the ease with which recordings can be made, stored, modified, and shared, substantially lowers the threshold for distributing information that may potentially be detrimental to someone's reputation. Whether it be through sharing original, modified, or synthetically generated "fake recordings," reputational harm may in some instances be inflicted with only a couple of clicks or taps. The fact that setting these processes in motion can frequently be done anonymously, potentially with few or no repercussions (reputational or otherwise) for the instigating party (cf. [26]), lowers this threshold even further.

Second, related to the Scaled Exposure Thesis, the reputational damage that can be inflicted through digital recordings in the digital ecosystem is larger than ever before because the recordings can be made available across contexts to a previously unthinkable number of people. Prior to the internet, for most individuals, one's social network was typically separated by different contexts and relatively small; the set of persons whose judgments mattered for one's life was also small. Now, a minor error of judgment or a slip of the tongue, recorded in a private context, can be broadcasted to one's entire social network and beyond, magnifying its impact tremendously.

Third, because one can be subjected to the judgment of so many people, the variety of standards by which one can potentially be judged is staggering. Even when people's moral worldviews are deeply wrong, their condemnation may, under certain conditions, nonetheless produce life-changing problems for those subjected to it. In this respect, the shift toward unbounded moralized judgment puts humans in what is in some sense a highly precarious situation: how can one maintain a positive or even neutral reputation in the eyes of a diverse audience of indefinitely many people over an indefinite amount of time?

Fourth, content from digital recordings can deliver substantial harm quickly, but it can also subtly but persistently erode one's reputation. In line with the Unpredictability of Modification and Use Thesis, modified audio or video snippets or stills may be used to shape a target person's public image (e.g., through memes). Of course, in principle, this may also work to one's benefit, for example when one is portrayed in a particularly heroic fashion, but it may also be used to foster an undesirable (and undeserved) stereotyped public image and cultivate negative associations.[34] One problem is that one cannot know beforehand in which direction one's reputation will be shifted; another problem is that in a world as diverse as ours, adulation is never universal, and even a small number of opponents can harm one's interests.

Fifth, and finally, the reputational harm that is inflicted may be enduring. In line with the Repeated Exposure Thesis, recorded actions have the potential to be observed over and over again, standing in the way of collective forgetfulness. Moreover, their contents can be used in a

[32] For many examples along these lines, presented in terms of online shaming, see Ronson [122] and Scheff & Schorr [123]; see Lim [124] for discussion of some cases in which bystanders in TikTok videos went viral and became targets for what we would characterize as moralized judgments; see also Solove [125], Citron [126], and Weissman [127,128].

[33] This is not to deny that reputations can also be positively affected by, and through, the digital ecosystem. Digital recordings of good deeds can be (and are) shared widely, and can also conceivably serve as personal reminders of one's moral achievements. However, as we take these positives to be culturally established, our focus here is on the pernicious but underappreciated ways in which reputations can be unjustly and irreparably damaged.

[34] Harris [129] discusses the possibility that deepfakes may cause significant harm to an individual just by influencing what people *associate* with that person.

variety of stigmatizing ways. This potential for lifelong stigmatization by (many) others, brought about by the long-term storage and sharing possibilities of the digital ecosystem, makes that it can be exceedingly difficult to recover from harm or to restore one's damaged reputation.[35] As Tunick [27] observes, "The punishment any particular individual inflicts cannot hope to be proportionate unless it is part of a coordinated response … the more wide-reaching and long-lasting the exposure of one's past misdeed is, the more likely nonlegal punishment will be disproportionate. Because of the coordination problem, there can be no assurance it will be measured or have an end" (p. 649). Whether there are cases in which such extensive and enduring reputational harm can be justified is an important question, but one that we will not be addressing here. The point we wish to establish is that, in making and letting other people make digital recordings of oneself, one increases the probability of incurring undue reputational harm (and all that follows from such harms). Likewise, when making recordings of others, regardless of one's intentions, one increases the probability that those recorded subjects will incur undue reputational harm.

We conclude, therefore, that the risk of incurring undue reputational damage presents a *pro tanto* reason for refraining from making digital recordings of oneself or others. When it comes to recording others, we propose that moral obligations related to that individual and to society generally supply a *pro tanto* reason against recording. In one's own case, prudential concerns alone produce such a reason, but in addition, moral considerations may also supply such a reason. For instance, Allen [28] (pp. 852–855) discusses how one might have second-order duties to protect one's own privacy, inasmuch as one has duties to care for or avoid harming others, and one's own compromised privacy may hamper one's ability to execute one's duties to others. Alternatively, if we take a consequentialist perspective, a damaged reputation may hinder an individual's ability to do as much good as they could do.

4.2. Tarnished self-perceptions

We have seen how the widespread sharing of digital recordings can affect one's social standing in ways that may be entirely misplaced because no wrongdoing has been committed or may be disproportionate to the wrongdoing that one has committed. But besides reputational harm, digital recordings also contribute to harm of a different kind. In this section, we focus on how practices involving the use of digital recordings may tarnish people's self-perceptions. Broadly speaking, we see two mechanisms through which this might happen, and we will discuss them in turn.

The first mechanism would be the way in which negative feelings toward oneself may be perpetuated when one is repeatedly reminded of past actions that one finds condemnable by one's own standards. How harshly one judges oneself exactly depends on the specifics of the situation and on one's dispositions (e.g., for neuroticism or self-criticism, see [29]), but it is a lived experience of many to feel shame, embarrassment, or guilt about certain past actions, to ruminate on one's past, and to engage in negative self-talk.[36] The risk, then, is that being confronted time and again with a growing set of recorded actions that showcase (what one views as) moral shortcomings will contribute to deterioration of one's self-image, potentially to the extent that it leads to a crisis of identity, loss of self-respect, or other mental health problems.[37] For people already prone to being overly self-critical, digital recording may worsen the situation; for others, recurrent encounters with digital recordings risk precipitating a condition of counterproductive self-criticism.

The second mechanism through which self-perceptions may be tarnished relates again to other people's opinions: namely, how beliefs held by others affect the way one views oneself. In particular, if people are subjected to moralized condemnation by those they respect or whose judgment they care about, viz. people who are in their "reference group" [30], the negative moralized judgments may be especially influential on one's self-evaluations. Certainly, there will be individual differences in how resilient people are in the face of negative sentiments from their social environment (cf. [31]). Few individuals, though, can endure a sustained barrage of negative judgments from people whose opinions they respect without being affected by it. At the least, when others judge that one has done something wrong, it can incline an individual to consider the *possibility* that they have acted wrongly. Thus, the risk is that if people—especially people from one's (inner and outer) social circle—are repeatedly invited or induced to pass judgment on one's shortcomings, by being repeatedly exposed to one's recorded past, the resulting negative sentiments from one's social environment will have an unduly negative effect on how one sees oneself.

Something that contributes to both of these mechanisms is digital recording's influence on the processes of forgetting and remembering. Plausibly, individual and collective forgetting constitute a very important feature of normative life, enabling societal re-integration of norm violators and influencing the development and maintenance of identities.[38] As Basu [32] argues, forgetting, as well as the possibility of being forgotten, plays an important role in the practice of shaping one's identity, in the sense of maturing, growing, and changing over time to the extent that we may no longer be the same people we were before ([32], p. 2). If, as we have been suggesting, digital recording practices are bringing about a shift toward unbounded moralized judgment, then they may have the effect of undermining identity-constructing processes, such as self-forgiveness and beneficial self-reinvention, by making it close to impossible to let go of past mistakes and to choose how one

[35] As others have previously argued, online shaming is often highly costly for the subject. To the extent that shaming constitutes a punishment for wrongdoing, it is frequently disproportionate. Billingham & Parr [130] argue that online public shaming often constitutes unwarranted punishment, inasmuch as it often fails to meet criteria of proportionality and accountability. (See also Aitchison & Meckled-Garcia [131] on how online public shaming as punishment can go wrong). Frye [120] argues that the technologies that support public shaming today encourage a form of shaming that is not reintegrative but rather disintegrative—even when shaming is done to promote prosocial ends (and even if those ends are good) rather than tending to bring norm violators back into the fold, online public shaming is an attack on a person's character that constitutes "an attempt to put social distance between the wrongdoer and other people" in a way that does not offer the possibility of redemption (p. 132).

[36] Of course, in moderation, such feelings and behaviors may facilitate learning from experience and moral improvement, but there is substantial evidence that excessive feelings and behaviors of those kinds can negatively affect one's self-image in a counterproductive way. Shame, for example, has been shown to have a large negative effect on self-esteem [132], where self-esteem is typically thought of as "an enduring state that relates to a fundamental sense of self-worth" that is "at the heart of human subjective well-being" ([133], p. 193–194; see also [134]). More generally, engaging excessively in self-criticism has been linked to depression, social anxiety, and other disorders (for a review, see [135]).

[37] It may be objected that in some instances and for some individuals, recordings of one's shortcomings may serve as a kind of crutch, reminding them not to engage in this kind of behavior again. Moreover, there certainly may also be recorded evidence of one's successes that keep individuals motivated. Even so, the general worry remains that, at least for certain groups of individuals, being repeatedly confronted with one's recorded shortcomings will negatively affect how they view themselves (as bad friends, inattentive parents, snide coworkers, poor community members, etc.), in a counterproductive way.

[38] On the topic of identity, see e.g., Tunick [27]. See also the discussion in Rini and Cohen [136] on the risk that people might use deepfakes to engage in "panoptic gaslighting"—causing an individual to become uncertain about whether they have committed particular actions, thereby corroding their sense of self. For comments on some of the benefits of forgetting, see Lundgren [137].

presents oneself to others in various contexts.[39]

Notice that what makes tarnished self-perception problematic from an ethical point of view is not necessarily that people incur damage to their self-image. One could say that viewing oneself in a negative light after engaging in morally condemnable behavior, and being judged by others for that behavior, is part and parcel of the human condition. Ethical hazards arise, though, when the moralized judgment is disproportionate to the wrongdoing, the moralized judgment is misplaced to begin with, or when the scaled and repeated exposure of one's shortcomings stand in the way of re-integration of norm violators. We conclude that these hazards present another *pro tanto* reason for refraining from making digital recordings of oneself or others.

4.3. Stifling moral behavior

The third and final hazard we wish to bring to the fore relates to the ways in which the increasing possibility in today's society of being (covertly) recorded may have a "chilling effect" with respect to people's morally relevant actions (cf. [20], p. 108).[40] To some extent, this hazard flows from the first two, in the sense that uncertainties with respect to the future use of recordings of oneself, including the risk of incurring undue harm to one's reputation and self-image, may lead to self-censorship and inhibition of one's actions ([33]; see also Marwick [16] on morally motivated harassment leading to self-censorship). Of course, certain forms of surveillance are employed precisely for the purpose of deterring people from engaging in morally *problematic* actions such as littering, cheating, or stealing (see, respectively, [34–36]), and presumably there will be circumstances where this has the desired effect.[41] What we want to highlight, however, is that there will also be circumstances in which individuals' anticipation of moralized judgment may lead them to morally worse decisions and actions. Here, we can draw on the literature on public surveillance, in which it is established that pervasive monitoring can have a range of negative effects. For example, it has been argued that extensive monitoring can undermine "an individual's ability to make choices about participation in social and political life" ([37], p. 1658), "[kill] free discourse and spontaneous utterances" ([163]), and "[threaten] not only to chill the expression of eccentric individuality, but also, gradually, to dampen the force of our aspirations to it" ([38], p. 1426).

In a similar vein, we hold that the increasing probability of being recorded at any moment by non-governmental parties—whether through commercially owned security cameras, street-facing home-monitoring systems, video-based productivity monitoring systems, screen recording software, dash cams, hidden pin cameras, or smartphones—can make people more hesitant to stand out. The risks of having one's actions, imagery, and words being taken out of context, stored indefinitely, and reused at will in various ways and at unexpected moments, can make people refrain from engaging in social activities, participating in public events, and speaking up about high stakes issues—both in public and in private.[42] Among these inhibited actions, there will be morally relevant ones, including participating in political activities such as public debates or protest marches, speaking up against injustices, engaging in the moral education of children, and giving expression to thought-provoking ideas in educational, academic, or healthcare contexts. The perceived risks of abuse and harassment already deters some people from holding public office or serving in leadership roles even at the local level ([39,40,41]). In contexts where topics like evolution, abortion, or race have been highly moralized [42], the threat of surreptitious classroom recording by students makes it potentially costly for professors to even broach those potentially-important-to-discuss topics.

As the hazards of unbounded moralized judgment increasingly become apparent, one might anticipate a "recorded bystander effect," whereby people hesitate to intervene in cases where wrongdoing is occurring. Although one might have expected that recording would encourage people to intervene to help, in many situations, what to do and how to do it is highly unclear and there will be disagreement on whether one has acted well.[43] Calls to emergency lines are now recorded in many countries and those recordings may be made public; interactions with law enforcement in many countries are recorded by dashcams and bodycams; in situations involving accidents or conflicts, bystanders often record what is happening with their phones. This could result in some people being less likely to call for emergency services or to attempt to help someone in need—because their attempt may fail or because they may (inadvertently) do something that others will condemn.

We recognize that societal pressures to conform to dominant social norms will not affect everyone to the same degree. Some people may be better positioned to withstand social backlash than others, and some may simply be willing to accept the risks of backlash so long as the potential rewards of engaging in those actions are high and salient enough. Still, we hold that if the social penalties of misstepping are high, and people will be subjected to moralized judgment from multiple (and changing) moralized worldviews by many people across contexts over an undefined period of time, then chilling effects are bound to occur. Here, our thinking aligns with Mill's remarks about social tyranny. Consider:

> [The] means of tyrannizing are not restricted to the acts which [society] may do by the hands of its political functionaries. Society can and does execute its own mandates; and if it issues wrong mandates instead of right, or any mandates at all in things with which it ought not to meddle, it practices a social tyranny more formidable than many kinds of political oppression, since, though not usually upheld by such extreme penalties, it leaves fewer means of escape, penetrating much more deeply into the details of life, and enslaving the soul itself. ([158], p. 4)

We have already seen the ways in which digital recordings, supported by the features of the digital ecosystem in which they are embedded, can be used by members of society to cause others undue harm of various kinds. Insofar as digital recordings are a means by which

[39] Consider, for example, the ways in which young individuals who are portrayed as a meme (e.g., "Side-Eying Cloe," "Ermahgerd," or "Success Kid" go through childhood being recognized and remembered as "that kid from that meme." We note that the experiences of these individuals vary, including over time; for some of their stories, see the "How I accidently became a meme" series on YouTube [138].

[40] Such effects are sometimes called "Panoptic" effects, after Bentham's utopian architecture for the threat of continuous individual supervision of inmates.

[41] The effectiveness of such systems often depends on whether appropriate processes are in place for determining how such surveillance footage will be shared and used in cases of alleged wrongdoing. Some people contend that police body cameras, for instance, have been ineffective in discouraging police abuse because police departments have too much control over what is done with body camera footage after incidents occur ([159]; [162]).

[42] Already a decade ago, some people were found to employ a "lowest common denominator approach" on social media, avoiding making statements if they believe that any possible viewer would find them problematic [139].

[43] Nguyen's [140] argument on how transparency can function as surveillance is also relevant here. In some circumstances, the right thing to do may be difficult to explain and defend to a highly diverse audience with varying backgrounds and forms of expertise. The modern recording ecosystem puts individuals in a position where they may face demands for justification from diverse, conflicting parties. When recorded individuals anticipate being called to account for each minute action they take, they may select actions that are morally worse but for which they can more easily articulate reasons to a diverse audience.

society can tyrannize, and insofar as we wish to create conditions in which members of society are not deterred from acting well, people have another *pro tanto* reason to refrain from being recorded themselves or recording others.

5. Emerging technologies and a look into the future

Thus far, we have considered the digital ecosystem as it currently exists and shown how characteristics of that ecosystem are drivers toward unbounded moralized judgment. However, there are several technologies on the horizon with the potential to make the hazards of unbounded moralized judgment substantially worse. Key to these technologies is that they rely on and build upon features of the current digital ecosystem that already contribute to the hazards we have identified. We will discuss three examples of such emerging technologies, but readily acknowledge that others may come to supplant them. Our aim in this section is to show that there are ongoing technological developments in a direction that may aggravate the trends we have concerns about.

One such technological development is augmented reality (AR). Facial recognition already makes it possible to uncover long-forgotten photos and videos of an individual from across the web—old home videos from childhood someone put on the internet; 20-year-old videos of the person out clubbing; footage of the person tripping, recorded by a doorbell camera; videos of oneself that one didn't know existed, posted to a pornography website. Imagine how all of this could be harnessed within augmented reality technologies like so-called smart glasses. Walking down the street while wearing one's AR glasses, the faces of strangers could be identified automatically, and a digital dossier of a stranger's top ten weirdest or worst or most shocking recorded moments auto-play as one passes by. The result would be an extreme sort of context collapse. It would be difficult to interact with other people while trying to overlook some insulting, thoughtless, or otherwise repellant thing they said or did ten years ago, which is quite fresh in one's mind because one's AR system just played it.[44]

A second potential technological development on the horizon is automated norm enforcement. One factor that constrains humans' capacity to generate moralized judgments and enforce norms is the fact that humans cannot view all the actions particular individuals perform on video every day. Given recent developments in image recognition and large multimodal models, we should anticipate the possibility that humans will attempt to harness AI systems to identify at least some types of (potential) norm violations on a massive scale. Already there have been efforts to use AI to automatically identify violence and aggression [43], inattention and undesired emotions in the classroom [44,45] and in the workplace [46], inattention while driving ([47], pp. 145–150), loitering [48], trespassing [49], and women leaving their hair uncovered in Iran [50]. In the future one might anticipate efforts to identify disrespectful actions, unpatriotic treatment of a national flag, inappropriate facial reactions (e.g., a failure to smile), insufficient deference to one's superiors in the workplace, women leaving their shoulders uncovered on Mormon college campuses, and so on. Minorities, including neurodivergent people, are particularly vulnerable in this scenario, but no one would be immune to the risks of being erroneously categorized or being punished for violating norms that some view as wrong.

There is a genuine question about whether any automated norm enforcement system that humans might attempt to deploy would actually be accomplishing its intended function.[45] For one thing, norm violation detection may require capacities that the systems do not have (e.g., an ability to reliably attribute emotion or intentions); for another, the error rate or particular distributions of types of errors can make AI systems nonfunctional—potentially without humans realizing it. One obvious source of such errors will be biases—it is known that AI systems trained using human-produced data tend to propagate or exacerbate existing societal biases ([51–53]; see also [54]). In addition, AI systems may produce new biases, not previously exhibited by humans [55–57]. An assortment of problems may result—e.g., efforts to promote one set of values may result in the system promoting a different set of values, and efforts to enforce norms may result in a disproportionate number of failures to identify norm violators from some groups and false allegations of norm violations for people from other groups.

A third technological development, intertwined with the first two, but also worth mentioning on its own, is recent progress in generative AI, particularly in the production of audio and video content that resembles some recognizable individual or group, but also in the form of chatbots. Generative AI involves a host of risks, many of which have been identified in existing overviews [58–60]. For the purpose of this article, the primary risks we wish to highlight have to do with the use of generative AI to facilitate the production of, or influence the spread of, negative moralized judgments. With regard to facilitating the production of negative moralized judgments, there is the issue of bad actors using digital recordings of their targets to create media that appear to depict individuals or groups as having properties or engaging in actions that will prompt consumers of the media to form negative moralized judgments of the target. For example, consider again individuals who create deepfake sex videos of people they know, depicting them in situations that many people consider degrading, with the aim of shaming them for perceived offenses—potentially causing the targets themselves as well as others online or offline to form negative moralized judgments about them.[46] Other examples include the use of deepfakes to depict politicians as saying something that the public will condemn, or deepfake depictions of minorities committing actions that inspire vitriol from other groups. Generative AI tools like online chatbots can also be used to influence the *spread* of negative moralized judgments, for political or other motives, such as when botnets employing generative AI are used to spread rumors about particular individuals or groups, such as that members of a particular ethnic group have engaged in a norm-violating action like animal sacrifice. One can imagine such chatbots employing personalization and tailoring techniques to generate an assortment of moralized messages that are catered to the recipients' views and more likely to be shared. These possibilities just scratch the surface of how generative AI may influence human practices of moralized judgment.

To be clear, these examples are not meant as a general critique of technology or innovation.[47] Rather, the purpose of the examples is to sketch what might happen if these technologies are brought to bear on a wide array of social interactions in a society already moving toward unbounded moralized judgment. Whether (or to what degree) these

[44] Cf. Tunick [27], who describes a similar scenario about using "smart" glasses on a plane. Relatedly, Lundgren [137] observes of AI-improved personal memory that it would put a strain on people's social relationships; the threat of AR combined with digital recording is that it would not only threaten one's existing social relationships but would also hamper the forming of new relationships—one would have access not only to recordings of events from one's own life but to recordings from other people's lives.

[45] As Raji et al. [141] observe, even critics of AI often assume that the AI systems "work" (in the sense of performing their intended function), but it is worth questioning whether this is so—in fact, Raji et al. allege, "Deployed AI systems often do not work" (p. 959).

[46] For recent reporting on the rising threat of AI-generated pornography, see e.g., Kraft [142] and Sang-Hung [143]. A 2023 study from a company called Security Hero reported that 98% of the 95,820 deepfake videos they found online in 2023 were pornographic [144]. On Kate Manne's analysis of misogyny, "If it feels like anything at all, it will tend to be righteous… It often feels to those in its grip like a moral pursuit, not a witch hunt. And it may pursue its targets not in the spirit of hating women but rather, of loving justice." (p. 20) ([145], quoted in [16]).

[47] Augmented reality applications, for example, may have relevant and justifiable applications in contexts of e-learning [146], cultural heritage [147], maintenance and repair [148], and so on.

prospective scenarios will develop in reality will depend on how we, as a society, decide to deal with recording technologies. We turn to mitigation strategies next.

6. A call for mitigation strategies

Our argument so far has been that each of the three hazards we identified gives individuals *pro tanto* reasons to refrain from being recorded or recording others. As such, the argument is directed at, and has implications for, individuals when they are making the decision whether to be recorded or record others. Such *pro tanto* reasons need not be decisive: there may be other, more weighty *pro tanto* reasons in favor of recording, such that, with respect to the relevant circumstances, one's overall, all-things-considered judgment will be in favor of recording. Moreover, there may be contexts in which there are societal considerations in favor of making recordings that outweigh the individual's considerations.[48] As scholars such as Walzer [166] and Nissenbaum [61] have argued, different contexts (e.g., healthcare, employment, air travel) may have different purposes or ends, which will sometimes have implications for how certain trade-offs are to be made (see also [62]). Such considerations affect the outcome of individual deliberation as well.

There is, however, a further dimension to our argument. For even with regard to those cases where there are substantial potential benefits to recording, and where *pro tanto* reasons against recording might be outweighed, societal questions arise about how to protect against the hazards of unbounded moralized judgment. That is, what can be done to guard against disproportional or otherwise undue harm in those cases where recordings are (justifiably) made? Obtaining the benefits of recording while avoiding the hazards in such cases requires altering our techno-social environment, whether by societal, legal, technical, or other means.

In this section, we will consider, in broad strokes, how norm-related, regulatory, and technical strategies could help mitigate the hazards arising from (the shift toward) unbounded moralized judgments. We conclude by calling for sustained further reflection as well as public and legislative debates about ways in which individuals in society can be protected from the hazards we have identified. Before continuing, however, two remarks are in order. The first pertains to the observation that, in practice, different types of strategies will often go hand in hand. For example, technical solutions may be required to enforce regulatory prohibitions, which subsequently affect societal norms. Likewise, technologies may change certain norms that subsequently become codified in law. We acknowledge that this is the case but choose to broadly distinguish types of strategies nonetheless for the sake of conceptual clarity.

Second, beyond the broad categories, we can think about introducing mitigation strategies at the following decision points: at the decision to record, during recording, while storing recordings, when sharing recordings, when analyzing recordings, or when experiencing (e.g., viewing or playing) recordings. At each of these points, there may be different approaches available, even within one type of strategy. For instance, the norms around making recordings may differ substantially from the norms around sharing those recordings, and the legal repercussions may differ substantially as well. In what follows, we will acknowledge this idea at several points of the discussion, but we note here that a full exploration of the differences between decision points and the different interventions they afford is a project for future work. Nonetheless, having this idea in the background will help to see how our suggestions here offer fruitful grounds for a potential framework of mitigation strategies to be developed in the future.

6.1. Norm-related mitigation strategies

The first kind of mitigation strategy concerns the various ways in which members of society can alter or create social and institutional norms to govern digital recording practices in a manner better suited to the (powerful capabilities of the) current digital ecosystem. As we have shown, today's digital ecosystem has evolved quite dramatically from the time when image, audio, and video recording technologies were first introduced into society. As such, members of society—individually, and as a constituency—have cause to reassess whether existing norms still apply, and whether they still advance their personal values as well as the values of society more broadly (cf. [62]). To the extent that the current norms are found wanting, there may be grounds for advocating norm change. For example, establishing and cultivating norms that foster forgiveness, respect, and compassion—even in the face of recorded mistakes and wrongdoings of the past—could conceivably go some distance toward reducing the impact that recordings can have on subjects' lives.[49] Likewise, norm change with respect to matters of justice and judgment itself (e.g., toward withholding judgment, or at least toward upholding a presumption of innocence), may to an extent lessen the damaging reputational effects.[50]

While the complex dynamics of norm change are still poorly understood (cf. [63]), it is known that norm change can occur at different levels (e.g., among friends, within a community, in national public arenas) and through different mechanisms [64].[51] Importantly, norms can change or evolve both through intentional and unintentional processes. In the context of digital recording, an example of an intentional change process can be observed in Korea, where activists have been attempting to change norms related to secret recordings of women (e.g., on public transit and in bathrooms), to inspire general intolerance of the making and consumption of such videos [65]. Likewise, in Western societies, some people are explicitly organizing portions of weddings, concerts, and other major events as camera- and phone-free—a policy facilitated via collecting devices at the door or placing devices in a locked pouch that prevents use [66,67].

Given how ubiquitous and pervasive digital recording is at present, and seeing how recording can fundamentally change how an activity is perceived and experienced, we suggest that decisions around recording should not be entirely spontaneous and unconsidered, nor unilateral. For that reason, we encourage reflection on, and explicit articulation of, recording-related norms in different contexts. More concretely, we call for society-wide debates on norms about whether to record, how to go

[48] Some may argue, for example, that people who hold positions in public office or in law enforcement have to accept being recorded more frequently and being subjected to a different level of public scrutiny of their (public role-related) actions and speech in order to meet their position's exacting standards of accountability. Likewise, some may argue that certain celebrities, in view of their position as public figures and role models, should accept being recorded more frequently. We do not engage with these ideas here. Rather, the point to notice is that the strategies we propose in this section are focused on mitigating *disproportionate* or otherwise *undue* harm, harm that may also—or in particular—befall celebrities and people in public office.

[49] In relation to forgetting, for example, Basu suggests that we need a norm of the following kind: "when we learn something about another due to an act that violated another's privacy, unless there are countervailing concerns of greater moral weight, we are required to forget what we've learned" ([32], p. 19).

[50] On this point, Tunick 27 recommends norm change to prioritize the use of alternative routes to obtaining justice without condemning the presumed guilty party to infamy: "encourage those with legitimate interests in free speech to find avenues of expression that are more sensitive to privacy interests" (p. 666).

[51] Scheff & Schorr [123], for example, reports of a case of learning from experience, describing the view of one victim of internet mockery: "Caitlin no longer finds Internet memes mocking individuals humorous and calls out those who do. 'Each one of those people is a real human being, a real person whose world imploded the day they found themselves to be a punch line on a giant stage,' she wrote. 'I know what it's like to be the person in that horrible photograph. I can't inflict such pain on someone else.'" (p. 232).

about recording (e.g., whether to share information with subjects about what information is being gathered), how to store and share recordings (e.g., perhaps images of a private family gathering should not be stored on publicly accessible websites), how different types of recordings may be analyzed or modified, or how to respond to different types of recordings (e.g., whether to laugh at cruel memes, or whether to base negative hiring decisions on recordings made when the applicant was a teenager).

As one possible reference point, consider Macnish, who in his work on "just surveillance" provides a set of criteria for when it is acceptable to record for the purpose of justice (i.e., recording acts of abuse for the purpose of holding someone accountable, etc.). He holds that "one should take into account the reason for the surveillance, the authority of the surveillant, whether or not there has been a declaration of intent, whether surveillance is an act of last resort, what is the likelihood of success of the operation and whether surveillance is a proportionate response" ([157], p. 142). Different contexts will likely require different criteria, but his work might function as a model for how to generate and structure such criteria. More theoretical work in this direction as well as empirical work on the topic of the near- and long-term consequences of recording in different contexts is needed to provide input to and guidelines for the debates.

At the institutional level, we call for institutions to implement clearer guidelines, where doing so would help protect against the hazards of unbounded moralized judgment. For example, we would welcome more explicitly recording-free zones in educational settings—to preserve or create bastions of freedom and creative expression, unhindered by unbounded moralized judgment. At the same time, in many contexts institutional prohibitions on recording will not be the answer. People can have good reasons to want to record themselves and their (consenting) loved ones. One problem, of course, is that recordings frequently capture the likeness of bystanders who have not consented to being recorded. In those instances, one person's freedom to record can infringe upon others' privacy rights. To deal with this type of scenario, we suggest exploration of how technologies might be used to resolve the value conflict. We will return to this idea in Section 6.3 but will first survey some ways in which legal avenues can be explored to mitigate the hazards we have identified.

6.2. Legal mitigation strategies

Depending in part on the outcomes of the societal debates, there may also be legal avenues to explore to instate better protections against the hazards of unbounded moralized judgments. Human rights are already protected by international law under the Universal Declaration of Human Rights (UDHR) and international and regional human rights treaties, and in many national and supranational legal contexts there are anti-discrimination laws that offer protection against the most blatant and egregious human rights transgressions that involve recordings. However, given the novelty of how recorded events are stored, shared, modified, and used, it may be that additional legal instruments—both soft law (recommendations, codes of conduct, standards) and hard law (e.g., international treaties)—are needed to be able to enforce certain norms and to offer remedial procedures to victims. For example, with regard to sharing videos, Takhshid [68] calls for a new privacy (tort) law of unwanted broadcasting, also noting that that others have called for regulating misuse of personal information and "objectification of crime spectators in light of widespread video recording of crime scenes" ([68], p. 144).

The possibility of using recordings for the creation of deepfake content has also prompted calls for regulation. Harris [69], for example, has called for a federal criminal statute in the U.S. to prohibit the publication of (pornographic) deepfakes. Several bills have been introduced since then (e.g., US HR3230 [155], US HR2395 [154], US HR5586 [156]) to "protect national security against the threats posed by deepfake technology and to provide legal recourse to victims of harmful deepfakes," but thus far, none have been enacted. Recent advancements with generative AI have renewed the worries about deepfakes, however, as evidenced by the 2023 Writers Guild of America strike, where actors demanded contractual protections against "performance cloning" or "digital clones," worrying that their likenesses could be owned and used in perpetuity by studios to create new "synthetic" performances based on source material from previous performances [70]. In addition, worries were expressed that "body scans" could make extras and body doubles obsolete. The additional protections that were negotiated in this industry could set a precedent for other industries to follow, either through contracts of their own or through industry standards and codes of conduct. Moreover, it could pave the way for more general copyright or other types of legislation to better protect the rights of recorded subjects.

In the European context, similar legal discourse is taking place about how to regulate deepfakes. Pertinent questions are being asked of the EU's current privacy and data protection regime, such as whether additional protections are warranted for the (mis)use of imagery of deceased individuals (e.g., [161]), whether additional labeling requirements are needed to notify users about deepfakes [71], and whether the current "notice and take down" regime places an undue burden on individuals themselves to "pursue their private images online, contact individual data controllers, and convince reluctant platforms to remove them, a tedious and often inefficient process" ([72], p. 3).

To an extent, these worries are addressed through recent European legislation, including the Digital Services Act and the Digital Markets Act, which, since coming into force in November 2022, impose due diligence obligations on online platforms, effectively requiring these service providers to moderate content and remove illegal content expeditiously (cf. [73]). Likewise, the European Union Artificial Intelligence Act (EU Regulation 2024/1689) also includes new provisions to ensure that content that is "artificially generated or manipulated" is disclosed as such (Art. 50). We welcome these developments but call for further research in this area to evaluate whether these provisions turn out to be effective measures to counter widespread malicious use of recorded and manipulated content and to examine their applicability in legal contexts outside the EU.

Finally, there are important legal (and societal) questions to be addressed in relation to the responsibilities of corporations to respect human rights. As many aspects of the digital ecosystem, including much of the infrastructure but also the many digital content sharing platforms, are owned and maintained by private corporations, it may be asked whether there are sufficient accountability structures in place to ensure that human rights are adequately protected (cf. [74]). In 2011, the UN Human Rights Council endorsed the Guiding Principles on Business and Human Rights (UNGPs), which outline the corporate responsibilities businesses have to protect human rights (including the right to privacy).[52] These principles are expressed as "societal expectations" instead of legal obligations and are often interpreted in the narrow sense of "do no harm" [75]. However, questions may be asked whether, under certain conditions, and within a limited scope, corporations should not also have positive obligations to ensure that rights are protected. A relevant development in this space is the European Corporate Sustainability Due Diligence Directive (CSDDD), which requires EU companies and non-EU companies operating in the EU to establish due diligence procedures to address potential adverse impacts of their actions—including practices such as incentivizing the making and sharing of digital content—on human rights. This directive may set a precedent for other countries as well to create a legal basis for more corporate accountability. We note that these are vexed issues, though, and call for careful scholarly scrutiny as well as public deliberation.

[52] For an excellent discussion on the implications of the UNGPs on corporate accountability, see Bernaz [149].

6.3. Technical mitigation strategies

Finally, technical mitigation strategies could help establish, maintain, and facilitate behavior in accordance with social and legal norms that mitigate the hazards of moralized judgment. The most basic way in which technology can do this is supplying people with options that help them avoid the hazards we have described.

With respect to recording, one example of adding a privacy-preserving option could be for video conferencing software to offer the possibility of automatically obscuring all identifying features in online video calls. Participants could be provided a lifelike or highly expressive avatar which allows for facial reactions and gaze direction, without revealing their appearance. Similarly, people could be offered options for voice alteration in order to protect against future misuses of their genuine voice. Such technical alterations to the socio-technical environment would respect people's freedom of choice and retain many of the benefits of today's video conferencing practices.[53]

Another technical mitigation strategy, one that could work in tandem with extending the set of available options, has to do with default settings.[54] For example, enabling background blurring by default may help prevent unsuspecting people at work or at local cafés being caught on camera just because someone in their vicinity decided to engage in a video call. Similar technical strategies could also benefit consumers who wish to take pictures of themselves in public by ensuring that front-facing cameras ("selfie cams") automatically blur individuals in the background, unless this is explicitly disabled in the camera app. These types of defaults could go some way to protecting the individuals who have not consented to having their identifiable likeness captured, while still respecting people's choice to record themselves.

In addition to setting defaults, technical alterations could be made to the digital ecosystem to provide people with additional information about the content they are posting and consuming. This could potentially help prevent some of the problems associated with moralized judgment based on deepfakes. Consider social media platforms. Instagram has already taken steps to classify draft posts and warn users if it considers the draft to be potentially bullying [76]. In connection to the legal obligations to disclose fabricated content (see above), confirmation questions could be added whenever someone is about to share unverified content (cf. [77]). On the viewer side, additional platform notifications could be associated with social media posts if the media content has not been verified as authentic by a source that viewers trust, signaling to viewers to take a more vigilant stance with regard to trusting (and further sharing) the content (see also the point about provenance below).[55]

Technical mitigation strategies could also be used to enforce prohibitions in special settings. One existing example can be found in the special formatting of some recorded Broadway shows that permit viewers to watch a scene featuring nudity, but not to pause, rewind, or fast forward [67]. Some platforms such as Signal also offer "view-once media" which are removed from the conversation thread once they have been viewed (also blocking efforts to make screenshots).[56] Along the same lines, there might be value in the creation of recording formats that are more difficult to transmit to public-facing platforms, or that can only be transmitted to specific parties when all recorded subjects have consented to those specific acts of sharing.

More generally, we would recommend directing resources into the development of technological methods for improving the control people have over their own digital likeness. Of particular relevance in this regard are watermarking technologies. Here, one could imagine innovations in watermarking technology (and their use), such that recording devices, including smartphones and screen-capturing software, could automatically detect if they were recording content that contained a signal indicating that it should not be captured. Such technologies could have a deterring effect on the practice of making clandestine recordings and screenshots of, e.g., Zoom calls, WhatsApp conversations, etc. that could then be used for other purposes.

Another important development in this space concerns "image provenance," techniques for determining if content has been modified, and if so, by which transformations (e.g., [78,79]). We find particularly interesting the ongoing efforts of the Coalition for Content Provenance and Authenticity (C2PA) to establish a protocol for automatically adding encrypted provenance information (e.g., photographer details, location, date) to digital content to be able to distinguish between authentic and modified content, and to make transparent the transformations a piece of content has gone through. Already, a number of companies, NGOs, and academics are working to adopt this standard through the Content Authenticity Initiative.[57] Given the worries about modified content, especially about individually or societally damaging deepfake material, we encourage other technology companies to consider adopting this protocol.

7. Conclusion

In this article we have argued that the interplay between features of digital recordings and the digital ecosystem in which they are embedded are facilitating a shift toward unbounded moralized judgment, which contributes to three significant but underappreciated hazards. We have shown how unbounded moralized judgment may harm reputations and tarnish self-perceptions, and how the risks of these harms may have a chilling effect on people's (moral) actions. We have done so with the dual aim of (1) showing that individuals have *pro tanto* reasons not to record themselves and others, and (2) asking questions of society more broadly about how to mitigate the hazards we have identified. In service of the latter part of our aim we have outlined three different kinds of mitigation strategies that can be pursued.

To obtain (and retain) the benefits of recording while avoiding the hazards of unbounded moralized judgment, we call for creative exploration of these mitigation strategies and for investment into (re)designing and (re)structuring aspects of society, including laws and norms, as well as the digital ecosystem, in ways that will help avoid or mitigate the hazards of unbounded moralized judgment. Given the immense technical, regulatory, and societal complexity of the digital ecosystem and its broader context, this work requires collaborative efforts from academics, engineers, and policymakers.

CRediT authorship contribution statement

B.A. Kamphorst: Writing – review & editing, Writing – original draft, Conceptualization. **E.R.H. O'Neill:** Writing – review & editing, Writing – original draft, Conceptualization.

[53] Of course, recorded subjects would need to be able to trust the companies providing such a service to not retain data about their actual face and voice; and they would need reasons to believe that future technologies would not make it possible to reverse the obscuring.

[54] Under the right conditions, defaults can have a powerful effect on behavior. See Jachimowicz et al. [150] for a meta-analysis on the effectiveness of default setting and a discussion of the limitations of this strategy.

[55] With this suggestion about verification labeling, we sound a word of caution. As one of the anonymous reviewers pointed out, empirical work on the "implied truth effect" suggests that only introducing labels for content that is by some measure considered *inauthentic* might under certain circumstances actually increase the perceived accuracy of inauthentic content that is (inadvertently) left unlabeled. For a discussion of this effect, see [151].

[56] It might also be worth considering if "view-once media" should be the default for sharing videos in chat conversations.

[57] See https://contentauthenticity.org/ (last visited on September 25th, 2024).

Data statement

Not applicable.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

None.

Acknowledgements

We are grateful to Janna van Grunsven, Björn Lundgren, and three anonymous reviewers for their constructive feedback on previous versions of this article and their helpful suggestions on how to improve it. Thanks to Helen Nissenbaum, Wybo Houkes, Patrik Hummel, Philippe Verreault-Julien, and Guido Löhr, and audiences at OZSW 2023 and IACAP 2023, for comments on earlier articulations of some of the ideas in this article. We thank Dawa Ometto for enlightening conversations on context and the metaphysics of events, and Nadia Bernaz for sharing her expertise on business and human rights to improve the section on legal mitigation strategies. Finally, we wish to acknowledge that this work was performed in the context of the research programme Ethics of Socially Disruptive Technologies (https://www.esdit.nl/), which is funded through the Gravitation programme of the Dutch Ministry of Education, Culture, and Science (https://www.government.nl/ministries/ministry-of-education-culture-and-science) and the Netherlands Organization for Scientific Research (https://www.nwo.nl/en/researchprogrammes/gravitation) (NWO Grant Number 024.004.031).

Data availability

No data was used for the research described in the article.

References

[1] M.D. Molina, What makes an internet meme a meme? Six essential characteristics, in: S. Josephson, J. Kelly, K. Smith (Eds.), Handbook of Visual Communication: Theory, Methods, and Media, Routledge, 2020, pp. 380–394.
[2] R. Millière, Deep learning and synthetic media, Synthese 200 (3) (2022) 1–27.
[3] R. Rini, Deepfakes and the epistemic backstop, Philosophers' Impr. 20 (24) (2020) 1–16.
[4] A. Javornik, B. Marder, J.B. Barhorst, G. McLean, Y. Rogers, P. Marshall, L. Warlop, "What lies behind the filter?" Uncovering the motivations for using augmented reality (AR) face filters on social media and their effect on well-being, Comput. Hum. Behav. 128 (2022) 107126.
[5] E. Maiberg, Inside the AI porn marketplace where everything and everyone is for sale, 404 Media (Aug. 22, 2023). https://www.404media.co/inside-the-ai-porn-marketplace-where-everything-and-everyone-is-for-sale/.
[6] N. Singer, Teen girls confront an epidemic of deepfake nudes in school, The New York Times (April 8, 2024). https://www.nytimes.com/2024/04/08/technology/deepfake-ai-nudes-westfield-high-school.html.
[7] N. Kristof, The online degradation of women and girls that we meet with a shrug, The New York Times (March 23, 2024). https://www.nytimes.com/2024/03/23/opinion/deepfake-sex-videos.html.
[8] R. Metz, Anyone can use this powerful facial-recognition tool—and that's a problem, CNN (May 4, 2021). https://www.cnn.com/2021/05/04/tech/pimeyes facial-recognition/index.html.
[9] R. Mac, C. Haskins, L. McDonald, Clearview's facial recognition app has been used by the justice department, ICE, Macy's, Walmart, and the NBA, BuzzFeed News (Feb 28, 2020). https://www.buzzfeednews.com/article/ryanmac/clearview-ai-fbi-ice-global-law-enforcement.
[10] S. Frenkel, This Russian program can find your face anywhere, BuzzFeed News (May 6, 2016). https://www.buzzfeednews.com/article/sheerafrenkel/this-russian-program-can-find-your-face-anywhere.
[11] Y. Welinder, A. Palmer, E. Selinger, J. Polonetsky, O. Tene, Face recognition, real-time identification, and beyond, in: E. Selinger, J. Polonetsky, O. Tene (Eds.), The Cambridge Handbook of Consumer Privacy, Cambridge University Press, 2018, pp. 102–124.
[12] T. Sharon, B.J. Koops, The ethics of inattention: Revitalising civil inattention as a privacy-protecting mechanism in public spaces, Ethics Inf. Technol. 23 (3) (2021) 331–343.
[13] S. Stich, The quest for the boundaries of morality, in: A. Zimmerman, K. Jones, M. Timmons (Eds.), The Routledge Handbook of Moral Epistemology, Routledge, 2019, pp. 15–37.
[14] E. O'Neill, E. Machery, The normative sense: What is universal? What varies? in: A. Zimmerman, K. Jones, M. Timmons (Eds.), The Routledge Handbook of Moral Epistemology, Routledge, 2019, pp. 38–56.
[15] A.P. Fiske, T.S. Rai, Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships, Cambridge University Press, Cambridge, 2014.
[16] A.E. Marwick, Morally motivated networked harassment as normative reinforcement, Social Media + Society 7 (2) (2021).
[17] K. Nahon, J. Hemsley, Going Viral, Polity, 2013.
[18] J. Jacquet, Is Shame Necessary?: New Uses for an Old Tool, Vintage, 2016.
[19] A.E. Marwick, d. boyd, I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience, New Media Soc. 13 (1) (2011) 114–133.
[20] D. Solove, Understanding Privacy, Harvard University Press, Cambridge, MA, 2008.
[21] G. Origgi, Trust and reputation, in: J. Simon (Ed.), The Routledge Handbook of Trust and Philosophy, Routledge, 2020, pp. 88–96.
[22] G. Origgi, Reputation: What It Is and Why It Matters, Princeton University Press, 2017.
[23] P. Barclay, Reputation, in: D.M. Buss (Ed.), The Handbook of Evolutionary Psychology: Volume I Foundations, second ed., John Wiley & Sons, 2015, pp. 810–828.
[24] J. Henrich, The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter, Princeton University Press, 2016.
[25] R.A. Smolla, Law of Defamation, second ed., Clark Boardman Co, New York, 2000.
[26] S. Levmore, The internet's anonymity problem, in: S. Levmore, M. Nussbaum (Eds.), The Offensive Internet: Speech, Privacy and Reputation, Harvard University Press, 2010, pp. 50–67.
[27] M. Tunick, Privacy and punishment, Soc. Theor. Pract. 39 (4) (2013) 643–668.
[28] A.L. Allen, An ethical duty to protect one's own information privacy, Ala. L. Rev. 64 (2013) 845–866.
[29] M.N. Servaas, H. Riese, R.J. Renken, J.B.C. Marsman, J. Lambregs, J. Ormel, A. Aleman, The effect of criticism on functional brain connectivity and associations with neuroticism, PLoS One 8 (7) (2013) e69606.
[30] C. Bicchieri, The Grammar of Society: The Nature and Dynamics of Social Norms, Cambridge University Press, Cambridge and New York, 2006.
[31] A. Oshio, K. Taku, M. Hirano, G. Saeed, Resilience and Big Five personality traits: A meta-analysis, Pers. Indiv. Differ. 127 (2018) 54–60.
[32] R. Basu, The importance of forgetting, Episteme 19 (4) (2022) 471–490.
[33] J. Kang, Information privacy in cyberspace transactions, Stanford Law Rev. 50 (4) (1998) 1193–1294.
[34] T. Nakamata, T. Abe, The effects of security camera, past littering, environment, and signboards on littering prevention, Shinrigaku Kenkyu: Jpn. J. Psychol. 87 (3) (2016) 219–228.
[35] A.M. Jansen, E. Giebels, T.J. Van Rompay, M. Junger, The influence of the presentation of camera surveillance on cheating and pro-social behavior, Front. Psychol. 9 (1937) (2018) 1–12.
[36] E.L. Piza, B.C. Welsh, D.P. Farrington, A.L. Thomas, CCTV surveillance for crime prevention: A 40-year systematic review with meta-analysis, Criminol. Publ. Pol. 18 (1) (2019) 135–159.
[37] P.M. Schwartz, Privacy and democracy in cyberspace, Vand. L. Rev. 52 (1999) 1609–1701.
[38] J.E. Cohen, Examined lives: Informational privacy and the subject as object, Stan. L. Rev. 52 (2000) 1373–1437.
[39] C.E. Anthony, T. Lee, J. Gottlieb, B. Rainwater, On the frontlines of today's cities: Trauma, challenges and solutions, National League of Cities (Nov 10, 2021). https://www.nlc.org/wp-content/uploads/2021/11/On-the-Frontlines-of-Todays-Cities-1.pdf.
[40] Brennan Center, Local election officials survey, March 10, 2022. https://www.brennancenter.org/our-work/research-reports/local-election-officials-survey-march-2022.
[41] D. Farmer, N. Lee, J. Day, High rates of harassment and threats may deter entry into local politics, Civic Pulse (Dec 14, 2022). https://www.civicpulse.org/post/high-rates-of-harassment-and-threats-may-deter-entry-into-local-politics.
[42] C. Flaherty, University tells professors to stay 'neutral' on abortion, Inside Higher Ed (Sept. 26, 2022). https://www.insidehighered.com/news/2022/09/27/university-tells-professors-stay-neutral-abortion.
[43] P. Sernani, N. Falcionelli, S. Tomassini, P. Contardo, A.F. Dragoni, Deep learning for automatic violence detection: Tests on the AIRTLab dataset, IEEE Access 9 (2021) 160580–160595.
[44] D. Gilbert, Facial recognition cameras are making sure kids in China pay attention in class, Vice (May 21, 2018). https://www.vice.com/en/article/wjbd7b/facial-recognition-cameras-china-class-attention.
[45] N. DiBerardino, L. Stark, (Anti)-intentional harms: The conceptual pitfalls of emotion AI in education, in: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023, June, pp. 1386–1395.
[46] K. Roemmich, F. Schaub, N. Andalibi, Emotion AI at work: Implications for workplace surveillance, emotional labor, and emotional privacy, in: Proceedings

of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–20.
[47] K. Levy, Data Driven: Truckers, Technology, and the New Workplace Surveillance, Princeton University Press, Princeton, NJ, 2023.
[48] M. Shanmughapriya, S. Gunasundari, S. Bharathy, Loitering detection in home surveillance system, in: 2022 10th International Conference on Emerging Trends in Engineering and Technology-Signal and Information Processing (ICETET-SIP-22), IEEE, April 2022, pp. 1–6.
[49] A. Zaman, B. Ren, X. Liu, Artificial intelligence-aided automated detection of railroad trespassing, Transport. Res. Rec. 2673 (7) (2019) 25–37.
[50] K. Johnson, Iran says face recognition will ID women breaking hijab laws, Wired (Jan 18, 2023). https://www.wired.com/story/iran-says-face-recognition-will-id-women-breaking-hijab-laws/.
[51] A. Caliskan, J.J. Bryson, A. Narayanan, Semantics derived automatically from language corpora contain human-like biases, Science 356 (6334) (2017) 183–186.
[52] V. Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor, St. Martin's Press, 2018.
[53] J. Buolamwini, Unmasking AI: My Mission to Protect What Is Human in a World of Machines, Random House, 2023.
[54] P. Slattery, A.K. Saeri, E.A. Grundy, J. Graham, M. Noetel, R. Uuk, J. Dao, S. Pour, S. Casper, N. Thompson, The AI risk repository: A comprehensive meta-review, database, and taxonomy of risks from artificial intelligence, arXiv preprint arXiv: 2408.12622 (2024).
[55] C. Stinson, Algorithms are not neutral: Bias in collaborative filtering, AI and Ethics 2 (4) (2022) 763–770.
[56] K. Takemoto, The moral machine experiment on large language models, R. Soc. Open Sci. 11 (2) (2024) 231393.
[57] E. Ferrara, The butterfly effect in artificial intelligence systems: Implications for AI bias and fairness, Machine Learning with Applications 15 (2024) 100525.
[58] V. Capraro, A. Lentsch, D. Acemoglu, S. Akgun, A. Akhmedova, E. Bilancini, J. F. Bonnefon, P. Brañas-Garza, L. Butera, K.M. Douglas, J.A. Everett, The impact of generative artificial intelligence on socioeconomic inequalities and policy making, PNAS Nexus 3 (6) (2024).
[59] T. Hagendorff, Mapping the ethics of generative AI: A comprehensive scoping review, Minds Mach. 34 (39) (2024) 1–27.
[60] L. Weidinger, J. Uesato, M. Rauh, C. Griffin, P.S. Huang, J. Mellor, A. Glaese, M. Cheng, B. Balle, A. Kasirzadeh, C. Biles, Taxonomy of risks posed by language models, in: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022, June, pp. 214–229.
[61] H. Nissenbaum, Privacy in Context, Stanford University Press, 2009.
[62] E. O'Neill, Contextual integrity as a general conceptual tool for evaluating technological change, Philosophy & Technology 35 (79) (2022) 1–25.
[63] G. Andrighetto, E. Vriens, A research agenda for the study of social norm change, Philosophical Transactions of the Royal Society A 380 (2227) (2022) 20200411.
[64] M.J. Gelfand, S. Gavrilets, N. Nunn, Norm dynamics: Interdisciplinary perspectives on social norm emergence, persistence, and change, Annu. Rev. Psychol. 75 (2024) 341–378.
[65] A. Gunia, "It breaks my heart." Confronting the traumatic impact of South Korea's spycam problem on women, Time (March 7, 2022). https://time.com/6154837/open-shutters-south-korea-spycam-molka.
[66] J. Morrisey, Your phone's on lockdown. Enjoy the show, The New York Times (Oct. 15, 2016). https://www.nytimes.com/2016/10/16/technology/your-phones-on-lockdown-enjoy-the-show.html.
[67] M. Stevens, With cameras on every phone, will Broadway's nude scenes survive? The New York Times (June 1, 2022). https://www.nytimes.com/2022/06/01/arts/broadway-nudity-phone-cameras.html.
[68] Z. Takhshid, Retrievable images on social media platforms: A call for a new privacy tort, Buff. L. Rev. 68 (2020) 139–197.
[69] D. Harris, Deepfakes: False pornography is here and the law cannot protect you, Duke L. & Tech. Rev. 17 (2018) 99–128.
[70] W. Bedingfield, Hollywood writers reached an AI deal that will rewrite history, Wired (Sept 27, 2023). https://www.wired.com/story/us-writers-strike-ai-provisions-precedents/.
[71] M.V. Huijstee, P.V. Boheemen, D. Das, L. Nierling, J. Jahnel, M. Karaboga, M. Fatun, Tackling deepfakes in European policy, Publications Office of the European Union, 2021. PE 690.039, https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf.
[72] K. Mania, Legal protection of revenge and deepfake porn victims in the European Union: Findings from a comparative legal study, Trauma Violence Abuse 25 (1) (2024) 117–129.
[73] A. Turillazzi, M. Taddeo, L. Floridi, F. Casolari, The digital services act: An analysis of its ethical, legal, and social implications, Law, Innovation and Technology 15 (1) (2023) 83–106.
[74] B.A. Kamphorst, A. Henschke, Public health measures and the rise of incidental surveillance: Considerations about private informational power and accountability, Ethics Inf. Technol. 25 (4) (2023) 1–14.
[75] F. Wettstein, Normativity, ethics, and the UN guiding principles on business and human rights: A critical assessment, J. Hum. Right. 14 (2) (2015) 162–182.
[76] E. Ravenscraft, Instagram's new anti-bullying nudges could actually work, Medium: OneZero (May 9, 2019). https://onezero.medium.com/instagrams-new-anti-bullying-nudges-could-actually-work-9811ef41b8cb.
[77] A. Collins, Forged authenticity: Governing deepfake risks, EPFL International Risk Governance Center (IRGC) (2019). https://infoscience.epfl.ch/record/273296C2PA. https://c2pa.org/.
[78] D. Moreira, A. Bharati, J. Brogan, A. Pinto, M. Parowski, K.W. Bowyer, P.J. Flynn, A. Rocha, W.J. Scheirer, Image provenance analysis at scale, IEEE Trans. Image Process. 27 (12) (2018) 6109–6123.
[79] D. Moreira, W. Theisen, W. Scheirer, A. Bharati, J. Brogan, A. Rocha, Image provenance analysis, in: Multimedia Forensics, Springer Singapore, Singapore, 2022, pp. 389–432.
[80] J. Hopster, The ethics of disruptive technologies: Towards a general framework. International Conference on Disruptive Technologies, Tech Ethics and Artificial Intelligence, Springer, Cham, 2021, September, pp. 133–144.
[81] J. Hopster, What are socially disruptive technologies? Technol. Soc. 67 (101750) (2021) 1–8.
[82] A. Bazin, H. Gray, The ontology of the photographic image, Film Q. 13 (4) (1960) 4–9.
[83] P. Atencia-Linares, M. Artiga, Deepfakes, shallow epistemic graves: On the epistemic robustness of photography and videos in the era of deepfakes, Synthese 200 (518) (2022) 1–22.
[84] S.D. Warren, L.D. Brandeis, The right to privacy, Harv. Law Rev. 4 (5) (1890) 193–220.
[85] J. Habgood-Coote, Deepfakes and the epistemic apocalypse, Synthese 201 (103) (2023) 1–23.
[86] K. Tenbarge, Found through Google, bought with Visa and Mastercard: Inside the deepfake porn economy, NBC (March 27, 2023). https://www.nbcnews.com/tech/internet/deepfake-porn-ai-mr-deep-fake-economy-google-visa-mastercard-download-rcna75071.
[87] FBI Public Service Announcement, Malicious actors manipulating photos and videos to create explicit content and sextortion schemes, June 5, 2023. https://www.ic3.gov/Media/Y2023/PSA230605.
[88] W. Hartzog, The fight to frame privacy, Mich. L. Rev. 111 (2013) 1021–1043.
[89] W. Hartzog, E. Selinger, Surveillance as loss of obscurity, Wash. & Lee L. Rev. 72 (2015) 1343–1387.
[90] A.L. Allen, Unpopular Privacy: What Must We Hide? Oxford University Press, 2011.
[91] H. Nissenbaum, Protecting privacy in an information age: The problem of privacy in public, Law Philos. 17 (1998) 559–596.
[92] B. Beal, The nonmoral conditions of moral cognition, Phil. Psychol. 34 (8) (2021) 1097–1124.
[93] E.E. Buchtel, Morality as fish: Defining morality as a prototype concept, Psychol. Inq. 34 (2) (2023) 80–85.
[94] S. Stich, The moral domain, in: K.J. Gray, J. Graham (Eds.), Atlas of Moral Psychology, The Guilford Press, 2018, pp. 547–555.
[95] E. Machery, R. Mallon, Evolution of morality, in: J. Doris (Ed.), The Moral Psychology Handbook, The Moral Psychology Research Group, Oxford University Press, 2010, pp. 3–46.
[96] E. Machery, Morality: A historical invention, in: K. Gray, J. Graham (Eds.), Atlas of Moral Psychology, Guildford Press, 2018, pp. 259–265.
[97] W. Sinnott-Armstrong, T. Wheatley, Are moral judgments unified? Phil. Psychol. 27 (4) (2014) 451–474.
[98] W.J. Brady, M.J. Crockett, J.J. Van Bavel, The MAD model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online, Perspect. Psychol. Sci. 15 (4) (2020) 978–1010.
[99] P. Rozin, The process of moralization, Psychol. Sci. 10 (3) (1999) 218–221.
[100] J.C. Wright, Morality as a regulator of divergence: Protecting against deviance while promoting diversity, Soc. Cognit. 39 (1) (2021) 81–98.
[101] L.J. Skitka, C.W. Bauman, E.G. Sargis, Moral conviction: Another contributor to attitude strength or something more? Journal of personality and social psychology 88 (6) (2005) 895.
[102] L.J. Skitka, J.H.F. Liu, Y. Yang, H. Chen, L. Liu, L. Xu, Exploring the cross-cultural generalizability and scope of morally motivated intolerance, Soc. Psychol. Personal. Sci. 4 (3) (2012) 324–331.
[103] J. Haidt, E. Rosenberg, H. Hom, Differentiating diversities: Moral diversity is not like other kinds, J. Appl. Soc. Psychol. 33 (1) (2003) 1–36.
[104] E. Turiel, The Development of Social Knowledge: Morality and Convention, Cambridge University Press, 1983.
[105] P.K. Stanford, The difference between ice cream and Nazis: Moral externalization and the evolution of human cooperation, Behavioral and Brain Sciences 41 (e95) (2018) 1–49.
[106] J. Haidt, The moral emotions, in: R.J. Davidson, K.R. Scherer, H.H. Goldsmith (Eds.), Handbook of Affective Sciences, Oxford University Press, 2003, pp. 852–870.
[107] N. Strohminger, S. Nichols, The essential moral self, Cognition 131 (1) (2014) 159–171.
[108] G.P. Goodwin, J.M. Darley, The psychology of meta-ethics: Exploring objectivism, Cognition 106 (3) (2008) 1339–1366.
[109] S. Levine, J. Rottman, T. Davis, E. O'Neill, S. Stich, E. Machery, Religious affiliation and conceptions of the moral domain, Soc. Cognit. 39 (1) (2021) 139–165.
[110] E. O'Neill, Kinds of norms, Philos. Compass 12 (5) (2017) e12416.
[111] V. Kumar, Moral judgment as a natural kind, Phil. Stud. 172 (2015) 2887–2910.
[112] V. Capraro, M. Perc, Mathematical foundations of moral preferences, Journal of the Royal Society Interface 18 (175) (2021) 20200880.
[113] J. Graham, J. Haidt, S. Koleva, M. Motyl, R. Iyer, S.P. Wojcik, P.H. Ditto, Moral foundations theory: The pragmatic validity of moral pluralism, in: Advances in Experimental Social Psychology, vol. 47, Academic Press, 2013, pp. 55–130.
[114] M. Atari, J. Haidt, J. Graham, S. Koleva, S.T. Stevens, M. Dehghani, Morality beyond the WEIRD: How the nomological network of morality varies across cultures, Journal of Personality and Social Psychology 125 (5) (2023) 1157–1188.

[115] O.S. Curry, D.A. Mullins, H. Whitehouse, Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies, Curr. Anthropol. 60 (1) (2019) 47–69.
[116] M. Alfano, M. Cheong, O.S. Curry, Moral universals: A machine-reading analysis of 256 societies, Heliyon 10 (6) (2024) e25940.
[117] K. Sterelny, The Pleistocene Social Contract: Culture and Cooperation in Human Evolution, Oxford University Press, 2021.
[118] P. Pettit, The Birth of ethics: Reconstructing the role and nature of morality. Berkeley Tanner Lectures, 2018.
[119] M. Tomasello, A Natural History of Human Morality, Harvard University Press, 2016.
[120] H. Frye, The technology of public shaming, Soc. Philos. Pol. 38 (2) (2021) 128–145.
[121] J.W. Patton, Protecting privacy in public? Surveillance technologies and the value of public places, Ethics Inf. Technol. 2 (2000) 181–187.
[122] J. Ronson, So You've Been Publicly Shamed, MacMillan, 2015.
[123] S. Scheff, M. Schorr, Shame Nation: The Global Epidemic of Online Hate, Sourcebooks (2017).
[124] C.-J. Lim, Someone else made you go viral on TikTok. Now what? Buzzfeed News (Feb 8, 2023). https://www.buzzfeednews.com/amphtml/clarissajanlim/viral-tiktok-consent-panopticontent.
[125] D.J. Solove, The Future of Reputation, Yale University Press, 2007.
[126] D.K. Citron, Hate Crimes in Cyberspace, Harvard University Press, Cambridge, 2014.
[127] J. Weissman, The Crowdsourced Panopticon: Conformity and Control on Social Media, Rowman & Littlefield Publishers, 2021.
[128] J. Weissman, P2P surveillance in the global village, Ethics Inf. Technol. 21 (1) (2019) 29–47.
[129] K.R. Harris, Video on demand: What deepfakes do and how they harm, Synthese 199 (5) (2021) 13373–13391.
[130] P. Billingham, T. Parr, Online public shaming: Virtues and vices, J. Soc. Philos. 51 (3) (2020) 371–390.
[131] G. Aitchison, S. Meckled-Garcia, Against online public shaming: Ethical problems with mass social media, Soc. Theor. Pract. 47 (1) (2021) 1–31.
[132] Y. Budiarto, A.F. Helmi, Shame and self-esteem: A meta-analysis, Eur. J. Psychol. 17 (2) (2021) 131–145.
[133] A. Salice, Self-esteem, social esteem, and pride, Emotion Review 12 (3) (2020) 193–205.
[134] N. Branden, The Six Pillars of Self-Esteem, Bantam, New York, NY, 1994.
[135] C.A. Loew, H. Schauenburg, U. Dinger, Self-criticism and psychotherapy outcome: A systematic review and meta-analysis, Clin. Psychol. Rev. 75 (101808) (2020) 1–19.
[136] R. Rini, L. Cohen, Deepfakes, deep harms, J. Ethics Soc. Philos. 22 (3) (2021) 143–161.
[137] B. Lundgren, Against AI-improved personal memory, in: J. Haltaufderheide, J. Hovemann, J. Vollmann (Eds.), Aging between Participation and Simulation: Ethical Dimensions of Socially Assistive Technologies in Elderly Care, De Gruyter, Berlin, 2020, pp. 223–233.
[138] BuzzFeed Video, Youtube playlist: "I accidently became a meme", Last updated on Dec 22, 2022. https://www.youtube.com/playlist?list=PL5vtqDuUM1DmriObn4hC4F73IiAoDTgKW.
[139] J. Vitak, The impact of context collapse and privacy on social network site disclosures, J. Broadcast. Electron. Media 56 (4) (2012) 451–470.
[140] C.T. Nguyen, Transparency is surveillance, Philos. Phenomenol. Res. 105 (2) (2022) 331–361.
[141] I.D. Raji, I.E. Kumar, A. Horowitz, A. Selbst, The fallacy of AI functionality, in: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022, June, pp. 959–972.
[142] Coralie Kraft, Trolls used her face to make fake porn, there was nothing she could do, The New York Times, July 31, 2024. https://www.nytimes.com/2024/07/31/magazine/sabrina-javellana-florida-politics-ai-porn.html.
[143] Choe Sang-Hung, South Korea, Misogyny has a new Weapon: Deepfake Sex Videos, The New York Times, Sept. 12, 2024. https://www.nytimes.com/2024/09/12/world/asia/south-korea-deepfake-videos.html.
[144] Security Hero, State of deepfakes: Realities, threats, and impact. https://www.homesecurityheroes.com/state-of-deepfakes, 2023 securityhero.io/state-of-deepfakes.
[145] K. Manne, Down Girl: The Logic of Misogyny, Oxford University Press, 2017.
[146] N.M. Alzahrani, Augmented reality: A systematic review of its benefits and challenges in e-learning contexts, Appl. Sci. 10 (16) (2020) 5660.
[147] I. Pedersen, N. Gale, P. Mirza-Babaei, S. Reid, More than meets the eye: The benefits of augmented reality and holographic displays for digital cultural heritage, Journal on Computing and Cultural Heritage (JOCCH) 10 (2) (2017) 1–15.
[148] S. Henderson, S. Feiner, Exploring the benefits of augmented reality documentation for maintenance and repair, IEEE Trans. Visual. Comput. Graph. 17 (10) (2010) 1355–1368.
[149] N. Bernaz, Conceptualizing corporate accountability in international law: Models for a business and human rights treaty, Human rights review 22 (1) (2021) 45–64.
[150] J.M. Jachimowicz, S. Duncan, E.U. Weber, E.J. Johnson, When and why defaults influence decisions: A meta-analysis of default effects, Behavioural Public Policy 3 (2) (2019) 159–186.
[151] G. Pennycook, A. Bear, E.T. Collins, D.G. Rand, The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings, Manag. Sci. 66 (11) (2020) 4944–4957.
[152] C. D'Orazio, K.K.R. Choo, An adversary model to evaluate DRM protection of video contents on iOS devices, Comput. Secur. 56 (2016) 94–110.
[153] D. Fallis, The epistemic threat of deepfakes, Philosophy & Technology 34 (2020) 623–643.
[154] H.R. 2395, 117th Congress: DEEP FAKES accountability act, GovTrack.us, https://www.govtrack.us/congress/bills/117/hr2395, 2021.
[155] H.R. 3230, 116th Congress: DEEP FAKES accountability act, Billtrack, https://www.billtrack50.com/BillDetail/1132741, 2019.
[156] H.R. 5586, 118th Congress: DEEPFAKES accountability act, GovTrack.us, https://www.govtrack.us/congress/bills/118/hr5586, 2023.
[157] K. Macnish, Just surveillance? Towards a normative theory of surveillance, Surveill. Soc. 12 (1) (2014) 142–153.
[158] J.S. Mill, On Liberty, Hacket Publishing, Indianapolis/Cambridge, 1859/1978.
[159] R. Moran, In Police We Trust vol. 62, Vill. L. Rev., 2017, p. 953.
[160] W. Sinnott-Armstrong, The disunity of morality, in: S. Matthew Liao (Ed.), Moral Brains, Oxford University Press, 2016, pp. 331–354.
[161] B.V.D. Sloot, Y. Wagensveld, Deepfakes: Regulatory challenges for the synthetic society, Computer Law & Security Review 46 (2022) 105716.
[162] E. Umansky, The failed promise of policy body cameras, N. Y. Times Mag. (Dec. 13, 2023). https://www.nytimes.com/2023/12/13/magazine/police-body-cameras-miguel-richards.html.
[163] United States v. White, 401 U.S. 745, 1971. https://supreme.justia.com/cases/federal/us/401/745/.
[164] N. Goodman. *Languages of Art: An Approach to a Theory of Symbols*, Hackett publishing, 1968.
[165] M. Kleiman-Weiner, R. Saxe, J.B. Tenenbaum, Learning a commonsense moral theory, Cognition 167 (2017) 107–123.
[166] M. Walzer, Spheres of Justice: A Defense of Pluralism and Equality, Basic Books, 1983.