Original Research Article

BIG DATA & SOCIETY

Big Data & Society
July–December: 1–12
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517221113772
journals.sagepub.com/home/bds
SAGE

Feeling fixes: Mess and emotion in algorithmic audits

Os Keyes[1] and Jeanie Austin[2]

Abstract

Efforts to address algorithmic harms have gathered particular steam over the last few years. One area of proposed opportunity is the notion of an "algorithmic audit," specifically an "internal audit," a process in which a system's developers evaluate its construction and likely consequences. These processes are broadly endorsed in theory—but how do they work in practice? In this paper, we conduct not only an audit but an autoethnography of our experiences doing so. Exploring the history and legacy of a facial recognition dataset, we find paradigmatic examples of algorithmic injustices. But we also find that the process of discovery is interwoven with questions of affect and infrastructural brittleness that internal audit processes fail to articulate. For auditing to not only address existing harms but avoid producing new ones in turn, we argue that these processes must attend to the "mess" of engaging with algorithmic systems in practice. Doing so not only reduces the risks of audit processes but—through a more nuanced consideration of the emotive parts of that mess—may enhance the benefits of a form of governance premised entirely on altering future practices.

Keywords

Algorithmic audit, facial recognition, affect theory, machine learning, trans studies, autoethnography

Audits, algorithms, and mess

As increased attention is given to the harms algorithmic systems produce, a range of proposals have been made for ameliorative strategies. One popular concept is that of the *algorithmic audit*; "assessments of the algorithm's negative impact on the rights and interests of stakeholders" (Brown et al., 2021: 2). Of particular interest for our purposes is the proposals for "internal audits"—that is, for audit processes designed to operate within organizations developing algorithmic systems (Raji et al., 2020; Rakova et al., 2021). Crucially, many of these proposals emphasize the need for audit(or)s to integrate broader perspectives in order to recognize the situated nature of both harms and the knowledge necessary to identify them ahead of time. Raji et al., for example, reference the "*essential* inclusion of independent domain experts and marginalized groups in the ethical review process" (Raji et al., 2020: 39; emphasis ours), while Hutchinson et al. identify, as one of their proposals for mitigating concerns about datasets underlying algorithms, the need to "consult diverse stakeholders" (Hutchinson et al., 2021: 564).

Although there are some concerns about the viability and limitations of these proposals (many of which are recognized by the proposers themselves), along with wider issues with "audit culture" in general (Ahmed, 2012; Seaver, 2019), our goal here is not to critique but to *complicate*. Specifically, we want to highlight the *messiness* of audits in practice, in contrast to the "meticulous and methodical" claims of audits' advocates (Raji et al., 2020: 33). We do so in order to productively raise questions about the ways in which audit models and processes often depend on a simplistic (neat) model of human engagement with audits, and the consequences of that engagement.

Writing about social inquiry more broadly, John Law uses "mess" to capture the ways in which the very phenomena we study and describe are frequently complex, vague, and incoherent (Law, 2004). As a consequence, methodologies which assume (or require) a rigid and simple phenomenon are ill-suited to meaningful inquiry. The result is less an idealized objective analysis, and more the reshaping and reprioritization of phenomena to fit the needs of a particular methodological lens or gaze.[1]

[1]Human-Centered Design & Engineering, University of Washington, Seattle, WA, USA
[2]San Francisco Public Library, San Francisco, CA, USA

Corresponding author:
Os Keyes, University of Washington, Human-Centered Design & Engineering, Campus Box 352315, Seattle, WA 98195, USA.
Email: okeyes@uw.edu

Law's description of mess captures what interests us here: the ways in which audit processes do or do not survive in the face of reality, and the experiential struggle of trying to make subject and object "fit." Neatly structured plans and processes provide stability, consistency, and certainty—but plans rarely survive contact with reality unscathed. The question then becomes *what kinds of messiness are obscured by these neat plans? What impact do they have on the viability, or consequences, of audits?*

One type of mess occurs as a result of the complexity of algorithmic systems themselves. Audit processes often assume that either a singular lifeworld or a singular object contains everything of interest in inquiring into algorithms. Raji et al., for example, explicitly describe their process as fitting an *internal* audit, one where the lifecycle of an algorithmic system—from idea to component parts to deployment—fits within a single organization. But as anthropologists of algorithms frequently remind us, bounding algorithmic systems are rarely as simple. Algorithms and datasets are frequently developed in a way that spans across boundaries, with a heavy dependence on third-party systems and their possibilities (Passi and Sengers, 2020).

Similarly, the network of relations that make up algorithmic systems and their consequences are rarely limited to a single private domain. Algorithmic systems' boundary-spanning nature frequently crosses not only between multiple private, for-profit companies but between different types of entities and organizations with dramatically different motivations, priorities, and frameworks of understanding. Prior case studies in ethical controversies in data science have demonstrated that it is often in precisely such dramatic jumps that a void of responsibility and predictability appears (Zimmer, 2018). Proposed audit processes, however, frequently assume not only a singular organization but a singular *for-profit* organization, at that, as the site of inquiry. As a consequence, the problems that designers of these processes seek to address are often oriented toward concerns within those environments—issues of trade secrets and how that might interfere with transparency, for example (Kitchin, 2017). Less-discussed are issues created when datasets are reused in truly unanticipated ways, outside of the direct control of the developers, often in ways that cross legal and ethical jurisdictions.

Finally—and in many ways, most importantly—there is the mess of emotion, feeling, and sensemaking bound up in auditing practices, as they are bound up in any human activity. As demonstrated adroitly by Sara Ahmed's writing on diversity work in higher education (Ahmed, 2012), the experience of providing oversight and undertaking audit work is frequently bound up in *experiences*. Undertaking this work may make one exhausted, or cause pain; it may induce anger at the injustices that have been revealed, or pride that they were corrected. Knowing the emotional involvement that comes with inquiry (Gilmore and Kenny, 2015; Kumar & Cavallaro, 2017), and the emotional involvement that comes with data (Kaziunas et al., 2017), it seems inevitable that feeling would appear in the experience of *making inquiry into data*.

Scholars examining social practices from feminist, disability studies, and critical race theorist perspectives have consistently highlighted emotion and feeling as simultaneously vital to understanding the experience of and consequences caused by a situation, and have noted that these are rarely considered in social inquiry and practice (Avgerou and McGrath, 2005; Smith, 1990). Normative western models of knowledge, and the disciplines that deploy them, have traditionally constructed their view and understanding of the world around a model of objectivity and rationality—one in which emotion, embodiment, and experience have no place (Code, 1991; Jaggar, 1989). As a consequence, the emotive experiences of both participants and researchers are discounted and actively erased from the ways we talk about, distribute, and articulate our work. Feeling is treated as an embarrassment—as a failure of objectivity—rather than a valid part of understanding (Gilmore and Kenny, 2015). Emotion is made *absent*—written through with (paradoxically, feelings) of purposeful avoidance (Scott, 2022).

This is a problem both because of the important role of emotion in sensemaking and social life, and the uneven distribution of what constitutes "rational" and "objective" ways of being and understanding. What constitutes rationality, and who can access it, is frequently coded and understood as male, white, able-bodied, heterosexual, and cisgender (Shotwell, 2011); not only does giving primacy to rationality risk producing monolithic ways of understanding, the polyvalent ways of knowing that are excluded are disproportionately deployed by marginalized populations—precisely those communities whose perspectives are most vital in understanding whether an algorithmic system perpetuates injustice.

Both the proposals for and executions of audit processes we have read are cautious about and cognizant of the harms of monolithic claims to rationality and truth, recognizing (for example) the importance of including a diverse range of experiential viewpoints in analyzing and understanding the system being audited. However, they often maintain, implicitly or explicitly, the *desire* for pure rationality, even if multiple perspectives are needed to approximate it. Frameworks have little or nothing to say about the *emotional* and *experiential* aspects of audits, which are (in western thought) regularly set aside from rationality and truth (Wilson, 2011: viii), and the possible harms (or benefits) tied up with those aspects (Code, 1991).

In an effort to break this pattern—the omission of mess in general, and emotion in particular—we have undertaken an audit with a twist. Rather than examining a system and removing our struggles, feelings, and difficulties from the resulting analysis, we decided to center them. To make the publication not simply about the results of our analysis, but about the difficulties—personal, practical, and structural—involved when any idealized process of analysis encounters reality. Fitting these desires, we have intentionally undertaken—and are reporting on—an audit of a machine learning dataset that untidily crosses between different contexts and organizations, and between our roles as abstract researchers and as situated, embodied, and human beings. We do this by documenting and reflecting on our audit of a facial recognition dataset, the "HRT Transgender Dataset." Although datasets are, quite clearly, not algorithms, the dependence of the latter on the former means that dataset evaluations are common to most audit proposals. Correspondingly, the experience of auditing datasets is directly relevant to (in fact, a subset of) auditing algorithms.

Methods of mess

Assembled by researchers at the University of North Carolina, Wilmington, led by Professor Karl Ricanek, the HRT Transgender Dataset consists of over a million images of 38 people, taken as frames from videos that had been uploaded to YouTube (Mahalingam and Ricanek, 2013). The dataset was promoted online on Ricanek's laboratory's website and made available to other researchers willing to fill out a consent form. A public source of data reprocessed into a lightly gatekept repository: there are hundreds of computer vision datasets just like this.

What made the HRT Dataset unique—justifying both its collection and its reuse—was the subject matter. The 38 subjects were transgender; their videos, "transition timelines," consisting of a series of videos (or a single video, featuring a compilation of photographs or edited together from multiple recordings) demonstrating and narrating the physical and other changes that occurred over a period of 12 or more months on hormone replacement therapy (HRT). Such videos are a common—indeed, almost stereotypical—form of trans-media production, providing both an opportunity for self-narration and monitoring and educational information about the experiential aspects of HRT (Horak, 2014). Many of these transition timelines were created in a social-temporal moment when the need for transgender in-group knowledge sharing was produced by the lack of information created *for*, much less easily accessible information created *by,* transgender people (Miller, 2019). In such an environment, social media platforms, including YouTube, have been important information sources for transition information because creators tended to highlight the embodied aspects of transition (Horak, 2014). The initial act of capturing the videos in the HRT Transgender Dataset removed them from this context.

Ricanek et al.'s purpose was neither self-narration nor education. Instead, their goal was to allow facial recognition systems to consistently track people despite the physiological changes HRT often produces. It was this that led them to creating, and releasing, the HRT Transgender Dataset. Information about the dataset was placed online in 2013, and Ricanek produced several journal articles that utilized it, along with an editorial describing the "novel challenges" to facial recognition systems created by HRT and other medical processes, and touting his dataset as the solution (Ricanek, 2013).

The public felt somewhat less positive about the dataset, and for good reason. Beginning in 2017, a range of journalists and commentators began publishing critical examinations of Ricanek's work. In particular, they highlighted disconnects around Ricanek's motivation and the question of consent. Transition timeline videos are frequently uploaded for in-community use; for "paying it forward." Ricanek's use of them was far from these altruistic and communal aims. Instead, he publicly expressed the motivating (and ludicrous) fear that terrorists might undergo hormone replacement therapy to sneak across the US border, evade matches with government-issued identification, or otherwise undertake hormone replacement therapy to nefarious ends (Vincent, 2017). These projected motivations mirror more general transphobic tropes—that transgender people are suspect, sneaky, and otherwise engaged in acts of trespass (between genders or borders) and subterfuge (Currah and Mulqueen, 2011; Fischer, 2019).[2] Further, although claiming that he had attempted to notify the videos' subjects, he undertook this "as a courtesy," rather than for consent purposes, explicitly including people regardless of whether they could be contacted. Pointing to the "current political climate," Ricanek claimed to have stopped distributing the dataset in 2014 and finally took down the page about it on his laboratory website following negative media coverage (Vincent, 2017).

Our choice to inquire into this dataset was initially motivated largely by curiosity–curiosity, and concern. Both the first and second authors of this paper are trans, and although neither of us are included in the dataset, both of us saw it as an exemplar of the violence that can occur when existing practices—the surveillance and over-examination of trans bodies and lives—begin to resonate with new technologies. We sought to understand the circumstances of the dataset's creation, use and redistribution, in order to map that violence and (possibly) ameliorate it.

Even early in the auditing process, it was impossible to neatly draw a line between our status as researchers and as (potential) subjects, or between our emotional reactions and insider knowledge and "objective," "rational" scientific gaze. In early encounters with the dataset, we found content created for in-group use, much of which incorporated vulnerability as a form of collective care, directed to other ends. The appropriation of the dataset into systems of policing and surveillance was in polar opposition to the content creators' motives of care—the appropriation of creators' images over time made them unwitting participants in state violence that directly targeted transgender people. The creation of the dataset also involved fixedness of identity—a "pre" and "post" transition self-defined primarily by medical intervention—that subjects may not have held over time. That compassion and a desire for the well-being of trangender people could so quickly be put to other ends left us disturbed, unable to distance our subjective experience from the purportedly objective role of the auditor.

This is not novel; as discussed above, we would argue that any such separation is ultimately artificial and harmful to the generation of understanding (Code, 1991; Wilson, 2011). Other scholars such as Ruth Pearce have written movingly about the tensions, feelings, and visceralities of being simultaneously subject and object (Pearce, 2020). Rather than attempt to elide this mess, we decided to use it; to make our *experience* of auditing part of the focus of our audit. Such an approach is increasingly common in the social sciences, appearing in methodological notions of (for example) affective autoethnography (Gherardi, 2019).

In structuring and approaching this, we were greatly influenced by the superb work of Cheryl Cooky, Jasmine R. Linabary, and Danielle J. Corple in modeling "feminist holistic reflexivity" in social media research (Cooky et al., 2018). Cooky et al.'s work melded feminist approaches to scholarship which recognize the role of the researcher within the process, with inquiries using "big data." This included not only conceptual commitment to recognizing and tracing paths through relations of power but working to put that commitment into practice through "individual journaling, collective responses to reflection questions, and recorded group discussions" (Cooky et al., 2018: 3). Encouraged by both this and existing work on the costs of research that suggests an (unsurprising) emotional burden that comes with this form of inquiry (Kumar and Cavallaro, 2018), we committed to regularly individually journaling each step of our investigation and difficulty, alongside collaborative meetings and exchanges to sympathize, articulate our experiences, or solve any sticking points.

Studying mess

Sourcing data

To unravel the story of the HRT Transgender Dataset, we began at the source—Ricanek's laboratory at the University of North Carolina, Wilmington (UNCW). Part of this felt akin to archeology: using the Wayback Machine and other tools to obtain access to earlier versions of the dataset's web presence and project page, tracing (in the sense used by Geiger and Ribes, 2011) how these changed over time, independently of and in response to the blowback the research team received. Simultaneously, we submitted requests to the University under the North Carolina public records law, seeking IRB submissions, team correspondence, and any other documentation around the research project. It is here that we first ran into the messiness of auditing in practice.

UNCW, the public records officer informed us, had not considered the research that led to the creation and distribution of the dataset as eligible for institutional review board review. Not only that, but the IRB itself had no record of Ricanek et al.'s work, having treated the project as lacking human protection concerns. Most pressingly for our efforts to understand the creation of the dataset, UNCW had no notes or records, and only fragmentary emails prior to 2014 due to a change in computing systems at that point.

Compounding this absence of data was an additional absence of engagement; we had reached out to Ricanek and his collaborators on the various papers documenting the dataset seeking their perspective on the project. The response was, almost uniformly, radio silence. Ricanek did not respond at all aside from forwarding the records request to an UNCW administrator, while Mahalingam deferred from the conversation, stating simply "I am no longer associated with UNCW or the [research] group. Dr Ricanek will be the best person to answer your questions." (personal communication, January 22, 2020). Mahalingam did not respond to an additional request for information related to the dataset creation and surrounding research, which occurred when she was associated.

Undeterred (well, slightly deterred), we dug into the data we *had* obtained: the material the public records officer was able to find. This consisted of approximately 90 emails, and associated attachments, spanning from 2013 to 2017. Digging into the emails revealed significant disparities between the researchers' public descriptions of events and what had actually occurred.

According to the researchers' interactions with journalists and dataset subjects, the dataset had been collected in 2011–2012 from transition timeline videos on YouTube that were marked as Creative Commons licensed, and so free for reuse and redistribution. Videos' creators (and therefore subjects) were contacted, and the dataset itself was not distributed to third parties—only links to the videos within it. Ricanek maintained to journalists and the public that the research team had stopped distributing even those links in 2014. Ricanek apologized to the participants who privately emailed the research team about never being contacted about their inclusion in the dataset, but framed this oversight as an unknown error in the consent protocol (Vincent, 2017).

On the surface, this was a plausible narrative of a perhaps naive but well-intentioned research team trying their best. The UNCW-provided documents contradicted almost every part of it.

Following the video links in question (discussed further in "Accounting for Data," below) did not turn up a single

Creative Commons licensed video; all those we could identify were provided under the standard YouTube license, which explicitly prohibited reusing and redistributing the content outside of YouTube as a platform at the time that the images were captured. Not only were there no records suggesting that participants had been contacted but one researcher on the project (Mahalingam) suggested quite the opposite. In an email to a third-party researcher, with Ricanek copied in, Mahalingam wrote that:

> "We are unable to let you use the images or the videos from this dataset in a public domain. This is due to the fact that these videos were…compiled to a dataset…without the subject's written concern." (Email from Gayathri Mahalingam, "RE: HRT Transgender Dataset–Use of stimuli by Project Implicit," 11 September 2015; bolding ours)

As the date of that email suggests, distribution did not stop in 2014, or 2015. Not only were we able to find examples of third-party researchers being given dataset access as late as the year the media furor broke, but the URL to the dataset was still accessible, without any password protection, in April 2021. This URL led not to a list of YouTube videos, as Ricanek had claimed to journalists, but to a Dropbox containing the video files in their entirety. At best, the researchers had been negligent with their statements to both subjects and media representatives: at worst, they had lied. Negligence alone cannot explain, for example, an exchange in 2015 in which Mahalingam emailed Ricanek asking if she could release "the cropped images from the Transgender dataset to the guy who has requested it. Technically, 13 videos have gone offline and we cannot share the images without the user's permission." Previous and future assurances to subjects and journalists pushed to the side, Ricanek responded simply:

> "That's fine…Just don't want to get in the habit of providing for everyone. Please do…" (Ricanek + Mahalingam, "RE: HRT dataset," 4 February 2015)

As discussed in our introduction, there are emotive and affective aspects to any task, including internal audits, and it is worth pausing to examine how emotion appeared at this stage of our work. Upon reflection, the first author approached the public records request, and conversations around it, in a mindset of suspicion and (frankly) paranoia; suspicion that records had not really been lost, paranoia that there might be an element of deception or economy with the *actualité* in the administrator's insistences. Queer paranoia and trans paranoia have a long history—indeed, Eve Sedgwick considers them foundational to queer theory as a discipline, and grounded in legitimate reasons (Sedgwick, 2003: 124). In this case, it reflected not only a generalized suspicion but the first author's specific experiences trying to access records at their own university during a fight for trans healthcare.

The second emotion—encountered while going through the records UNCW provided—was *anger*. Anger at the researchers; anger at the researchers' past actions; anger at a distance. This, too, is familiar: as Hil Malatino writes, "trans rage" is a common response to a world in which "we must rely on relationships with people and institutions that interpret us as subhuman, or at the very least misrecognize us so profoundly that the 'I' conjured in interaction barely resembles the 'I' we understand ourselves to be" (Malatino, 2019: 126). This anger is a response to harm and can be harmful itself, but as Malatino also notes (drawing on the work of Audre Lorde and Maria Lugones) it can serve as a catalyst, and as a source of energy and analysis. In this case, both understandings were present; there was a cost to the experiences that induced such anger, but the anger simultaneously provided the fuel to take a closer, and more exacting look at the data we were studying.

Tracing data

We next turned to tracing the afterlives of this dataset. From the presence of a Database Release Agreement on the UNCW lab's archived website, we were already aware that the data were possibly being reused outside of the university. The records we received contained extensive documentation on the reuse that had occurred, including the completed Agreements and associated metadata about the reusers, their institutions, and their projects. Our next task became using that metadata, along with online traces, to identify how (and to where) the dataset had spread. From there, we hoped to get further information from reusers about how they had in turn used and handled the dataset.

We were struck by how broadly the dataset had spread, including into disciplines with their own histories of transphobia and to scholars who likely lacked the background knowledge needed to critically contextualize the creation of the dataset. The records contained 16 requests for the dataset—all approved—from 15 institutions spanning seven countries. Requests came from multiple disciplines, including not only computer vision but also psychology, marketing, and business studies, and ran the gamut from doctoral dissertations to undergraduate capstone projects. The sheer breadth of legal jurisdictions demonstrates the difficulty of scaling formal processes of tracing and auditing. Nevertheless, we informally reached out to many of the reusers, seeking to understand how they had made use of, secured, and understood the HRT dataset.

We did not take making contact with the researchers lightly. We were in conversation with the IRB at the University of Washington about whether or not this contact constituted research, and were in frequent conversation with one another about the ethical boundaries and stakes that we had created in choosing to conduct an external audit. We held the IRB's decision that this was not formal research alongside our own feelings of responsibility, and many of our journal entries reflected (and reflected on) this tension.

There is no one universally "correct" answer to the problems here. The one we settled on is that reaching out to researchers was necessary to discharge our broader responsibilities, both as researchers and within our non-professional communities. We decided that reaching out to dataset reusers—offering participants in our study precisely the knowledge of their enrollment that Ricanek's original subjects were denied—was the right thing to do.

As with the initial UNCW researchers, many reusers simply did not respond. Those few who did confirmed that, despite withdrawing the dataset, Ricanek had never contacted them to seek the destruction of their copies or even let them know that the dataset was being scrutinized. Some explicitly stated that they had never passed copies on; others, more vaguely, stated that they could not remember. But we found evidence suggesting that multiple reusers —including one of those who had "forgotten" doing so—had passed the data on. One researcher who sought access as a PhD student later became a professor at a different university and gave the dataset to his students, in turn (Spielmann and Stern, 2021).[3] A slightly more complex example, demonstrating how data travels not only between researchers but between universities and countries, is a computer vision project from Norway that made use of the HRT dataset. None of the project authors were logged as requesting access to the dataset. But all of them had, in an earlier project, collaborated with a researcher in India who was—and had presumably redistributed it to his collaborators.[4]

This part of the work felt more complicated, emotionally, than the last Some of the same feelings and responses made an appearance; there was the same anger at each unpleasant discovery, and the same frustration and helplessness provoked by each refusal by reusers to answer our questions (or: decision to do so disingenuously, or deceptively). Yet there was also joy—or perhaps exultation—with every instance of puncturing that deception, of being vindicated in our suspicion. More broadly, there was a sense of solidarity. The way that we went about this stage of the work was particularly collaborative, with the authors following traces while on the phone with each other. It became a collective activity, one where we got to share, in real time, each others' discoveries, empathize on any uncovered horrors, and revel in each others' joy at new discoveries. There were also moments of unexpected solidarity *with some reusers*. The discovery of (one) trans reuser, who had chosen not to use the dataset due to precisely the same discomfort that motivated us to investigate this project, made for a moment of shared understanding (across continents and disciplines), and a spark of connection and community for people who often exist, in the academy, in isolation and fragmentation (Pitcher, 2018).

These silver linings were somewhat undercut by repeatedly encountering images from the dataset in the papers we were examining. Computer vision researchers often include example images from the dataset their project uses in publications about it—examples of the dataset broadly, or specimen images that "failed" or "passed" their analysis. In the case of projects reliant on the HRT dataset, this included images of transgender people that, despite the withdrawal of the dataset by Ricanek et al. and ongoing questions and suspicions about their provenance, were effectively fixed in the public eye thanks to academic conventions around the sanctity and immutability of published works. We found individuals crystallized in identities they no longer held, their "before" pictures situated next to the "after," biometric analysis stripping them from selfhood and intention, transphobic histories of freaks on display, problems to be solved, lurking only barely under the surface. We returned again and again to the permanency of publication and the impossibility that transgender content creators might first know that their images circulated in these publications, and then find a way to reclaim or remove the images by navigating the tangle of academic publishing. Even Ricanek, in an email exchange with someone asking to have their images removed from publications, acknowledged the difficulty of doing this, claiming that there was not a way to remove the images from "bootlegged copies." Here, we identified with the video creators at a remove, not because of alignment between our lives but through the aching familiarity of what it is to lose agency, control, recognition, and volition of the (transgender) self (Keyes, 2020).

Accounting for data

At this point, we have discussed both our efforts to understand the creation and distribution of the HRT dataset, and its legacy outside of the original research team. What more is there? The answer is the *participants*—or subjects, if we want to echo the HRT Dataset's model of capture. The answer is our ethical duties as researchers, and as trans people, and the tensions involved in attempting to realize and align them.

Reading through the UNC Wilmington documentation demonstrated that (researcher assurances to journalists notwithstanding) there were reasons to be suspicious of claims that dataset subjects had been asked for their videos' inclusion, or even notified after the fact. As discussed above ("Sourcing Data"), UNCW's records revealed complaints from participants about a lack of notification, along with emails from research team members implying a uniform absence of consent. Finding this raised new questions and concerns for us: what, if any, was *our* duty to dataset subjects? Should we, for example, notify them about their inclusion, at a bare minimum? What responsibilities did we have—what opportunities might there be—to offer some recourse?

Under a traditional ethic of volunteerism, we have no duties here: we had not been involved in collecting the data, so had no responsibility to notify the subjects of it. But as this example illustrates, an ethic of volunteerism is both limited and highly questionable; it legitimizes the leaving of recognized injustices uncorrected. Based on both our professional ethico-political orientations and personal commitments, we felt an ethic of *care* was more appropriate, in the sense articulated by Eva Feder Kittay; an ethic in which "the needs of another call forth a moral obligation on our part when we are in a special position vis-à-vis that other to meet those needs" (Kittay, 2019: 62; see also Linabary and Corple, 2019). Casey Rebecca Johnson gives the example of a drowning child, one unknown to the observer; an ethic of care states that "I have not volunteered to care for her or protect her from harm, but I nonetheless should help. I have noticed her, am physically proximate to her, I can swim and so can attempt to help her without undue risk—I am in a special position to meet her needs. I have an obligation to help her because I'm well placed to do so, and because she badly needs help" (Johnson, 2020: 678–9). Datasets are not drownings—but nevertheless, we felt that having noticed the harm that a denial of knowledge and agency represented, and being in a "special position" as a result of our access to the dataset and its documentation, we, too, had an obligation. We decided to contact the dataset participants and notify them of their presence in the dataset, offering whatever information or conversation we could.

Doing so required reidentifying subjects, relying solely on a set of YouTube links almost a decade old. Our expectation was that this would be somewhere between difficult and impossible—but in practice, this was not the case. Even in instances where videos had subsequently been marked private or taken offline entirely, tools like the Wayback Machine and YouTube itself allowed us to reconstruct 29 videos and their contributors' profiles, along with (in 17 cases) names, email addresses or links to other social media accounts. That this reidentification was even possible suggests Ricanek erred in not making this project subject to IRB supervision: UNCW explicitly requires a formal submission if "a subject [can] be individually identified by any data, information, or specimens you obtain" (UNC Wilmington Research Integrity Office, 2021), even if that data is initially public, and serves as a reminder of the sort of "intimate link that remains between an image's 'by-product' and its provenance, even after the data has been 'processed' and 'pulverised' as Big Data" (Thylstrup, 2019: 6).

Our experience of gathering this data was marked by multiple forms of discomfort, at multiple levels. The first site of concern was simply how much data we could gather, despite the amount of time that had passed, and how this served as a reminder that "the internet is forever"; that the traces we leave behind in digital platforms and datasets can always be reconstructed in ways that catch the tracee (and tracer) by surprise (Draz, 2018; Keyes, 2021). At an intellectual level, this is something we were well aware of—academia and the world are replete with examples of and commentaries on this phenomenon. But there is something about finding oneself cast *in* this play of actors that is visceral and destabilizing in a way a purely-cognitive awareness is not.

Just as complex were our feelings around the videos, and their authors. By default, YouTube "autoplays" videos upon loading—which meant that in following links to try to find authors, we had to watch their videos. For the first author, this provoked a roiling mix of identification and disidentification; of community and insecurity. They found themself measuring their own value against the subject of the videos, judging themself for the gaps between the idealized, linear transition timeline and narrative that such works represent and their own experience (Haimson et al., 2020). As they wrote in their journal, "Her skin is gorgeous….I'd look awful. I could never look like her. I'd be such a disappointment. She's so confident. I wish I could be that confident. I wish I wasn't so scared." Engaging with other people, or their traces, is not just a formal, rational, professional task but also literally an "engagement with the other"; a confrontation with and recognition of different ways of being, and a reflexive comparison between those ways and one's own (Scott, 2019; Strauss, 2017). In a society with highly regularized and "wounded" (Westbrook, 2020) forms of trans life, it is easy to find not only an abundance of others' data, but a lack in one's self.

The second author was left with the precarity of it all. For many of the content creators, especially in the timeframe in which they were creating, there was an inherent risk in creating content. These are risks that transgender people in public know too well—of ridicule, of vitriol, of violence—but these are risks that can be assumed and accounted for and held in what are often acts of self- and collective-desire. Malatino has pointed to these networks as a public-private "where we access forms of preservative love withheld in the popular domain" (Malatino, 2020: 67). Even though the second author had steered clear of videos like these during their own physical transition—an act facilitated by the privilege of existing in proximity to other trans people, and one often informed by a reluctance to engage in spaces that created intense forms of intimacy—there was a feeling of disgust at this level of exposure. At so many steps, the content creators had been denigrated, pushed out, their intimacy laid open to the world, lacking the recognition of an agentive self behind it.

This was (dis)quieting: the only way to have sidestepped this risk, one the creators could not have predicted, would have been silence. So many documents around transition are simultaneous to share information and to claim a stake in the self (Horak, 2014). In the creation of the HRT Transgender Dataset, that was severed. Would the content creators have chosen to make their bodies, thoughts, pain, pleasure, and vulnerabilities available if they knew those images would be put to this use? This is an unanswerable question. What they will do with the knowledge that their images have been brought into the dataset, and in some instances endure separately from their consent or awareness, is not.

We decided to notify them, emailing those dataset participants we could identify with information about ourselves and our research project, what led us to the audit, and a notification that their images were included in the dataset. We informed individuals that the dataset was supposedly no longer available. We included anticipated questions and any answers to them that we could provide, and invited each content creator to contact us for further information, if they would like to do so.

None of the seven people we notified responded.

We still sit with that silence, and all that it might imply.

Discussion

In our analysis above, we have narrated and explored our experiences conducting an audit of the "HRT Transgender Dataset," with a particular focus on the difficulties encountered and our affective experience in engaging with the dataset, its creators, and its reusers. Although we successfully unearthed an array of information—much concerning—about the dataset's production, distribution, and use, we also encountered a wide array of "mess," both in our efforts to acquire information and in our efforts to grapple with our relation to the dataset and the figures around it. It is to the implications of this mess that we now turn.

Material messes

As discussed in the introduction to this paper, internal audits are a commonly-proposed remedy to concerns about algorithmic systems' negative consequences, particularly their *unintended* consequences. Audits are often portrayed as simple, linear, and regularized—as a matter of following formal processes and procedures. But as our own experience shows, and as other researchers have demonstrated about auditing work more generally (Star and Strauss, 1999), there is nothing simple about it.

Audits are inherently retrospective, and so ultimately rely on information about their target's creation and use being logged, stored, and made available in some form to auditors. But systems fail; people leave things unwritten, or they leave, full stop. Even looking at the traces around a dataset collected by a large, public-sector organization with longstanding processes for records access, we encountered infrastructural failures and data losses. One would expect these issues to be more, rather than less, common in auditing algorithmic systems more broadly, since many of them (including some of the most widely-used) are deployed by private-sector entities without those same traditions, who exist in a wide range of sizes with a correspondingly wide range of infrastructural and institutional stability.

Discussing institutional stability brings us to the second type of "material messiness" we found; messy *boundaries*. Technological workflows (including algorithmic ones) often exceed an individual team or organization, and regularly exceed an individual legal or policy jurisdiction. An algorithm may be developed in Canada using a dataset originally collected by a university in France, as reformatted by a third-party contractor in Belarus. Indeed, the outsourcing of stages of dataset creation, reworking, and labeling is *de rigueur*, to the point where entire companies exist solely to serve as this kind of third-party contractor (Keyes, 2020).

Contrary, then, to the assumptions of audit approaches that aim to establish conventions within one organization or nation, unpicking the historic trajectories (let alone future consequences) of an algorithmic system frequently involves many entities, in many locations (Bellanova et al., 2021). Absent any set of responsibilities—formalized or otherwise—understood between, as well as within, such actors, third-party (or perhaps, vicarious) entities involved are in no obligation to participate. This can be exacerbated when those entities' institutional memory is simply lost; when, for example, a postdoctoral researcher departs and takes with them all of their work not written down.

Our point here is not to say that audits are pointless, but rather that, as currently conceptualized, they are *toothless.* This toothlessness stems from a failure to adequately grasp just how many moving parts are implicated and imbricated in the work of making audits, well, work. Notwithstanding broader normative concerns about the adequacy of audits even, in theory, proposals for what they might look like in practice need to go a lot further than prescribing *auditors'* actions. They need to carry obligations for other actors—reusers, consumers, intermediaries—and an understanding of precisely how much institutional machinery must function to enable the transparency and memory audits require.

Affective presences

The second focus of our analysis was the affective and emotional experience of undertaking this kind of audit. How did it feel to actively seek out and delve into possible harms and their real or imagined repercussions? There is no single answer: at different points (and to different authors), this project was experienced as featuring anything from unexpected recognition and community to fear and anger.

Speaking broadly, however, the experience was largely a negative, unpleasant one. We felt rage, horror, and helplessness; we felt vicariously vulnerable and violated, and righteously enraged at the injustices we were observing.

But that we experienced and expressed these feelings marks, in some respects, our "outsiderness" to internal audit processes within technology companies—and not in ways that reflect well on those companies. We were advantaged by being in an environment where, at least in theory, critical inquiry and analysis is given primacy rather than treated in a purely instrumental way that "suppresse[s] critique that pose[s] a threat to productivity" (Su et al., 2021: 5) We were further advantaged in being a degree (or two, or ten) removed from the researchers whose work we were analyzing, rather than embedded in or a formal part of the team that had constructed the dataset.

In contrast, relations *within* formal organizations—the relations involved in the notion of an "internal audit"—are often heavily regulated. "[E]motional management" and "affective regulation," often manifesting as "a disposition not to feel" (Jones et al., 2019: 87), is a commonplace phenomenon within society broadly and high-scrutiny professional organizations in particular (Saifer and Dacin, 2021). This includes technology companies, where feelings are often seen as playing (at best) second fiddle to ideals of rationality, and often treated explicitly as a hindrance (Su et al., 2021). Were we conducting this analysis in such an environment, we would be expected to engage in precisely that regulation. Furthermore, we would be doing so in a situation where we would be expected to have ongoing, and otherwise "productive," relations with the very researchers whose work we were investigating. Perhaps we are overly cynical—perhaps the reader of this paper is far more generous than the authors—but we cannot imagine being able to simply look the original researchers in the eye, quash our feelings about their work, and carry on with the day-to-day practices of collaboration. Affects, as Clare Hemmings notes, "do not only draw us together, whatever our intentions; they also force us apart" (Hemmings, 2012: 153).

Unable to engage in such regulation, we would become the problem; the "killjoy," as Ahmed puts it, the person who is informed that "in assuming we have a problem, you are the problem" (Ahmed, 2012: 179). As Ahmed's choice of wording ("we") communicates, affective regulation is often experienced and enforced in a differential fashion, one that acts to preserve the comfort of those with power and problematizes the expressions of already-marginalized people (Hall, 2007; Jones et al., 2019; Srinivasan, 2018). In other words, the very people identified by Raji et al. as "essential" to the audit process, precisely due to their relation to injustices and marginalization, are simultaneously both more likely to experience a degree of torque and more likely to be punished for expressing it. We can see (and in the first author's case, as a former worker in the technology industry, have experienced) precisely this dynamic occur (Amrute, 2019).

Were we instead to *succeed* in regulating and quashing signs of our experiences, we would risk not only incompletely and inadequately presenting the depth of harm involved in the HRT Dataset's creation and distribution, but visceral and psychic costs to ourselves, to boot. Researchers have consistently documented the misery and cost of internalizing negative experiences, without avenues through which to express them (Saifer and Dacin, 2021). Indeed, a refusal of a community to acknowledge the harms occurring within it is often experienced as secondary harm (Walker, 2006), and a refusal to acknowledge the affective components is an "affective injustice" (Srinivasan, 2018). In an environment where affective experiences are silenced, it is difficult to avoid the feeling that audit processes might, at best, avoid immiserating marginalized users only through immiserating marginalized employees. To a certain degree, this already happens at other points in the process of algorithmic development—take, for example, the exposure of (often underpaid, racialized, and precariously employed) social media moderators and algorithmic "labelers" to extreme and traumatic content. The dynamic there, as here, is to redistribute misery on to them for the sake of the user; to treat them as sin eater, and a necessary victim to avoid further victims in turn (Gray and Suri, 2019, Roberts, 2019). Rather than simply integrate "diverse stakeholders" into audit processes, then, audit processes that seek to avoid emiserating outcomes must take care to examine and attend to the consequences *for those involved*, including emotive consequences, and acknowledge the ways in which "epistemic inclusions may be just as pernicious as epistemic exclusions" (Pohlhaus, 2020: 234).

It would be easy to interpret this as a call for "listening"—a call for audit process designers, and those executing them, to examine not just the algorithm, but how the audit process enables them (or does not) to "better [understand] the different anxieties that various participants [are] experiencing" (Gould, 2009: 331). Enabling such understanding is certainly part of our desires, as cautious as we are about the dangers of an (un-critical) empathy (Hemmings, 2012: 153; Ortega, 2006). But the weaponization of care in otherwise-conventional environments is a familiar phenomenon (Srivastava, 2006), including in the technology sector. In 2021 alone, we have seen companies such as Google and Accenture (on behalf of Facebook) use putatively-"caring" practices (offers of therapy, promises of "wellness") not to address harms but to minimize and redirect them (Gupta and Tulshyan, 2021; Satariano and Isaac, 2021).

Instead, we suspect that the work of making internal audit processes that avoid creating harm of its own will require more widespread changes to the organization in question; efforts to ensure "not only caring individuals but the active support of caring members through organizational goals, systems, strategies, and values" (Lawrence and Maitlis, 2012: 644). Addressing injustice requires "both thinking carefully and caring thoughtfully" (Hall, 2007: 92), and while these practices make themselves known through individual actions, they are enabled (or not) by wider, structural processes in turn (Liedtka, 1996).

Affective possibilities

We opened our analysis by describing it as a complication, rather than a critique; an effort to open up foreclosed problems and possibilities. Thus far, the discussion has focused on the problems, and even problematized some *solutions*. But we do want to emphasize that we find hope, here, too, and reasons to pursue even imperfect ameliorative efforts.

To identify that hope, we should take a step back; what is the goal of an audit? In the design of the examples we have pointed to, the goal seems to be to minimize the deployment of harmful algorithmic systems (and so minimize the harms). This goal is laudable but pursued largely through examining and changing the algorithmic systems themselves.

We would argue that a far more productive approach is not to fix harmful algorithms, but to "un-invent" them (MacKenzie, 1993); to rework organizations and processes to the point where a harmful algorithm is not just not deployed, but never made; to the point where the idea of making an algorithm harmful in a particular way is simply unintelligible to the developers. Doing so requires not only intervening in the algorithm, but promoting change in the knowledge and practices of the algorithmic designers—in creating space for unlearning (sometimes subtle, but always pernicious) ignorances and biases. Engagement with questions of feelings can be a powerful tool in doing that.

Decisions—including decisions in the design of algorithmic systems—are hardly made in a uniformly explicit, contextless, and rational way. Instead, they are often dependent on tacit and implicit knowledge and (situated) understandings, and heavily laced with emotive heft. This includes the biases and ignorances that concern researchers theorizing about algorithmic audits; they are not necessarily explicit by default, and so must be *made* explicit to be addressed.

In *Knowing Otherwise*, Alexis Shotwell (2011) makes the case that "The implicit may be visible at sites of a certain rupture in habitual activity…moments of strong emotion or unpremeditated reaction," terming this an "affective shock" (Shotwell, 2011: xvi-xx). In making this argument, Shotwell aligns strongly with critical feminist perspectives on pedagogy, which emphasize the potential for "discomfort…as a possible critical impetus for change, and for thinking and knowing differently" (Chadwick, 2020: 5), and thus the necessity for a "politics of discomfort" (Applebaum, 2017: 682).

What does this mean in the context of algorithmic development? It means that attending to feelings—not only the sense of injustice experienced by marginalized actors within these environments but the sense of discomfort or defensiveness this is likely to prompt in developers themselves—making that discomfort, and the clash between assumptions and beliefs that it represents, explicit. Correspondingly, it renders it into a form that can be discussed, reflected on, and addressed.

While we are cautious, once again, about how the possibilities here are shaped by broader structural aspects of these companies, we are nevertheless hopeful that attending to questions of feeling in algorithmic audits has the potential to not only reduce the harm that comes to parties to those audits, but, further, make explicit the underlying conditions that led to the algorithm in question's possible harms, and offer the possibility of changing those very conditions.

Conclusion

Algorithmic injustices matter, and so too does attending to them. Proposals to do so through new processes in development environments make bold claims—but, as our analysis has demonstrated, their very process-oriented *nature* elides complications of infrastructure, feeling, and experience that threaten to undermine the entire edifice. Addressing these complications requires going far beyond new processes and demands revisiting the very dynamics of power and work the technology sector depends on. Demands, indeed, change in the very structures of the sector.

This does not call for despair, but instead, for hope; for the hope that through precisely this revisiting, we can create a world not only of less harmful algorithms but of more helpful developers. Such a world requires not practices for developing invulnerable software, but practices for allowing vulnerable *people*.

Acknowledgments

The authors are grateful to Michael Zimmer for his involvement in the formative stages of this project, Kat Walsh for insight on the pettifoggery of YouTube licensing, Hil Malatino for his consistently-resonant work on trans affects, Alainna Brennan Brown for editing, Margret Wander, Shannon Hackett and Claire and Margaret Hopkins for their moral support, and—of course—each other.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Microsoft Research through an Ada Lovelace Fellowship.

ORCID iD

Os Keyes https://orcid.org/0000-0001-5196-609X

Notes

1. As Sara Ahmed succinctly notes, "we have been taught to tidy our texts, not to reveal the struggle we have in getting somewhere" (Ahmed, 2016, p.13).
2. Broader discussions of "the spoofer" as a motivating figure in facial recognition research can be found in (Grünenberg, 2019).
3. This researcher later told us that he could not remember, "off the top of [his] head," sharing the dataset at all.
4. To avoid replicating some of the distributional harms these papers commit, we are not citing works that contain images from the HRT dataset; for alternative (and brilliantly thoughtful) approaches to the same issue, the reader should see Cagle (2021)'s notion of "ethical ekphrasis."

References

Ahmed S (2012) *On Being Included*. Durham: Duke University Press.

Ahmed S (2016) *Living a Feminist Life*. Durham: Duke University Press.

Amrute S (2019) Of techno-ethics and techno-affects. *Feminist Review* 123(1): 56–73.

Applebaum B (2017) Comforting discomfort as complicity: White fragility and the pursuit of invulnerability. *Hypatia* 32(4): 862–875.

Avgerou C and McGrath K (2005) Rationalities and emotions in IS innovation. In: Howcroft D and Trauth EM (eds) *Handbook of Critical Information Systems Research*. Cheltenham: Edward Elgar, 299–334.

Bellanova R, et al. (2021) Toward a critique of algorithmic violence. *International Political Sociology* 15(1): 121–150.

Brown S, Davidovic J and Hasan A (2021) The algorithm audit: Scoring the algorithms that score US. *Big Data & Society* 8: 1.

Cagle LE (2021) The ethics of researching unethical images: A story of trying to do good research without doing bad things. *Computers and Composition* 61: 1–14.

Chadwick R (2020) On the politics of discomfort. *Feminist Theory* 22(4): 556–574.

Code L (1991) *What Can She Know?* Ithaca, NY: Cornell University Press.

Cooky C, Linabary JR and Corple DJ (2018) Navigating big data dilemmas: Feminist holistic reflexivity in social media research. *Big Data & Society* 5: 2.

Currah P and Mulqueen T (2011) Securitizing gender: Identity, biometrics, and transgender bodies at the airport. *Social Research* 78(2): 557–582.

Draz M (2018) Burning it in? Nietzsche, gender, and externalized memory. *Feminist Philosophy Quarterly* 4(2): 1–21.

Fischer M (2019) *Terrorizing Gender: Transgender Visibility and the Surveillance Practices of the US Security State*. Lincoln: University of Nebraska Press.

Geiger RS and Ribes D (2011) Trace ethnography: Following coordination through documentary practices. *2011 44th Hawaii international conference on system sciences*.

Gherardi S (2019) Theorizing affective ethnography for organization studies. *Organization* 26(6): 741–760.

Gilmore S and Kenny K (2015) Work-worlds colliding: Self-reflexivity, power and emotion in organizational ethnography. *Human Relations* 68(1): 55–78.

Gould DB (2009) *Moving Politics*. Chicago, IL: University of Chicago Press.

Gray ML and Suri S (2019) *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston: Eamon Dolan Books.

Grünenberg K (2019) Wearing someone else's face: Biometric technologies, anti-spoofing and the fear of the unknown. *Ethnos* 87(2): 223–240.

Gupta AH and Tulshyan R (2021) 'You're the Problem': When They Spoke Up About Misconduct, They Were Offered Mental Health Services. *The New York Times*, 28 May, 21.

Haimson OL, et al. (2020) Trans time: Safety, privacy, and content warnings on a transgender-specific social media site. *Proceedings of the ACM on Human-Computer Interaction* 4(CSCW2): 1–27.

Hall C (2007) Recognizing the passion in deliberation: Toward a more democratic theory of deliberative democracy. *Hypatia* 22(4): 81–95.

Hemmings C (2012) Affective solidarity: Feminist reflexivity and political transformation. *Feminist Theory* 13(2): 147–161.

Horak L (2014) Trans on YouTube: Intimacy, visibility, temporality. *Transgender Studies Quarterly* 1(4): 572–585.

Hutchinson B, et al. (2021) Towards accountability for machine learning datasets: Practices from software engineering and infrastructure. *Proceedings of the 2021 Conference on Fairness, Accountability, and Transparency (FAT* '21)*, pp 560–575. New York: Association for Computing Machinery.

Jaggar AM (1989) Love and knowledge: Emotion in feminist epistemology. *Inquiry* 32(2): 151–176.

Johnson CR (2020) Epistemic vulnerability. *International Journal of Philosophical Studies* 28(5): 677–691.

Jones K, Schroeter F and Schroeter L (2019) Mind-making, affective regulation, and resistance. *Australasian Philosophical Review* 3(1): 86–89.

Kaziunas E et al. (2017) Caring through data: Attending to the social and emotional experiences of health datafication. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp.2260–2272. New York: Association for Computing Machinery.

Keyes O (2020) Automating autism: Disability, discourse, and artificial intelligence. *The Journal of Sociotechnical Critique* 1(1): 1–31.

Keyes O (2021) (Mis)gendering. In: Thylstrup NB (eds) *Uncertain Archives*. Cambridge: MIT Press, 339–346.

Kitchin R (2017) Thinking critically about and researching algorithms. *Information, Communication & Society* 20(1): 14–29.

Kittay EF (2019) *Love's Labor: Essays on Women, Equality and Dependency*. Abingdon: Routledge.

Kumar S and Cavallaro L (2018) Researcher self-care in emotionally demanding research: A proposed conceptual framework. *Qualitative Health Research* 28(4): 648–658.

Law J (2004) *After Method: Mess in Social Science Research*. Abingdon: Routledge.

Lawrence TB and Maitlis S (2012) Care and possibility: Enacting an ethic of care through narrative practice. *Academy of Management Review* 37(4): 641–663.

Liedtka JM (1996) Feminist morality and competitive reality: A role for an ethic of care? *Business Ethics Quarterly* 6(2): 179–200.

Linabary JR and Corple DJ (2019) Privacy for whom?: A feminist intervention in online research practice. *Information, Communication & Society* 22(10): 1447–1463.

MacKenzie DA (1993) *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance*. Cambridge: MIT Press.

Mahalingam G and Ricanek K (2013) Is the eye region more reliable than the face? A preliminary study of face-based recognition on a transgender dataset. In: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)* (pp. 1–7). IEEE.

Malatino H (2019) Tough breaks: Trans rage and the cultivation of resilience. *Hypatia* 34(1): 121–140.

Malatino H (2020) *Trans Care*. Minneapolis, MN: University of Minnesota Press.

Miller JF (2019) YouTube as a site of counternarratives to transnormativity. *Journal of Homosexuality* 66(6): 815–837.

Ortega M (2006) Being lovingly, knowingly ignorant: White feminism and women of color. *Hypatia* 21(3): 56–74.

Passi S and Sengers P (2020) Making data science systems work. *Big Data & Society* 7: 2.

Pearce R (2020) A methodology for the marginalised: Surviving oppression and traumatic fieldwork in the neoliberal academy. *Sociology* 54(4): 806–824.

Pitcher EN (2018) *Being and Becoming Professionally Other: Identities, Voices, and Experiences of US Trans* Academics*. Bern: Peter Lang.

Pohlhaus G (2020) Epistemic agency under oppression. *Philosophical Papers* 49(2): 233–251.

Raji ID, et al. (2020) Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. pp 33–44. New York: Association for Computing Machinery.

Rakova B, et al. (2021) Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction* 5(CSCW1): 1–23.

Ricanek K (2013) The next biometric challenge: Medical alterations. *Computer* 46(9): 94–96.

Roberts ST (2019) *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven, CT: Yale University Press.

Saifer A and Dacin MT (2021) Data and organization studies: Aesthetics, emotions, discourse and our everyday encounters with data. *Organization Studies* 43(4): 623–636.

Satariano A and Isaac M (2021) The Silent Partner Cleaning Up Facebook for $500 Million a Year. *The New York Times*, 31 August, 21.

Scott S (2019) *The Social Life of Nothing: Silence, Invisibility and Emptiness in Tales of Lost Experience*. Abingdon: Routledge.

Scott S (2022) Social nothingness: A phenomenological investigation. *European Journal of Social Theory* 25(2): 197–216.

Seaver N (2019) Knowing algorithms. In: Vertesi J and Ribes D (eds) *digitalSTS*. Princeton: Princeton University Press, 412–422.

Sedgwick EK (2003) *Touching Feeling*. Durham: Duke University Press.

Shotwell A (2011) *Knowing Otherwise*. University Park: Pennsylvania State University Press.

Smith DE (1990) *The Conceptual Practices of Power: A Feminist Sociology of Knowledge*. Toronto: University of Toronto Press.

Spielmann J and Stern C (2021) Gender transition shapes perceived sexual orientation. *Self and Identity* 20(4): 463–477.

Srinivasan A (2018) The aptness of anger. *Journal of Political Philosophy* 26(2): 123–144.

Srivastava S (2006) Tears, fears and careers: Anti-racism and emotion in social movement organizations. *Canadian Journal of Sociology* 31(1): 55–90.

Star SL and Strauss A (1999) Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer Supported Cooperative Work (CSCW* 8(1): 9–30.

Strauss A (2017) *Mirrors & Masks: The Search for Identity*. Abingdon: Routledge.

Su NM, Lazar A and Irani L (2021) Critical affects: Tech work emotions amidst the techlash. *Proceedings of the ACM on Human-Computer Interaction* 5(CSCW1): 1–27.

Thylstrup NB (2019) Data out of place: Toxic traces and the politics of recycling. *Big Data & Society* 6: 2.

UNC Wilmington Research Integrity Office (2021) Human Subjects Research Protection. Available at: https://uncw.edu/sparc/integrity/irb.html#IRBDecisionCharts (accessed 26 August 2021)

Vincent J (2017) Transgender YouTubers had their videos grabbed to train facial recognition software. *The Verge*. Available at https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset (accessed 12 May 2021)

Walker MU (2006) *Moral Repair: Reconstructing Moral Relations After Wrongdoing*. Cambridge: Cambridge University Press.

Westbrook L (2020) *Unlivable Lives: Violence and Identity in Transgender Activism*. Berkeley: University of California Press.

Wilson EA (2011) *Affect and Artificial Intelligence*. Seattle, WA: University of Washington Press.

Zimmer M (2018) Addressing conceptual gaps in big data research ethics: An application of contextual integrity. *Social Media +Society* 4: 2.