CATALYST
feminism, theory, technoscience

CRITICAL PERSPECTIVES

Digital Apprehensions: Policing, Child Pornography, and the Algorithmic Management of Innocence

Mitali Thakor
Northwestern University
mitali@northwestern.edu

Abstract

This paper examines how racial bias plays out in the algorithmic detection and classification of child pornography images by police and non-police actors, including computer scientists and content moderators at technology companies. I employ the concept of "digital apprehension" to argue that the way programmers and moderators understand child pornography is always already oriented toward the arrest of virtual offenders. As image recognition algorithms—which are used to detect both victims and offenders—rely on white skin tones as defaults, with reduced accuracy with nonwhite skin tones, innocence is brought into close proximity with whiteness. Perceptive and algorithmic bias bear high stakes for police-work, as policing expands into a global network carried by the issue of child pornography.

Thakor, M. (2018). Digital Apprehensions: Policing, Child Pornography, and the Algorithmic Management of Innocence**,** *Catalyst: Feminism, Theory, Technoscience*, *4*(1), 1-16. http://www.catalystjournal.org | ISSN: 2380-3312

Introduction

I peered over Linda's shoulder as she sat in her office cubicle looking at her computer. Her screen displayed an array of photos of young boys in different positions and clothing, their heads and faces blurred. On the top left of her screen was the image she had originally uploaded (a white boy in a blue baseball cap) to see if any matches of similar photos could be found in her database. Many of the resultant images clearly showed the same child but in different clothes and poses. She pointed at her screen and looked up at me. "See? Take a look at that."

Linda is an analyst with the Child Victim Identification Program at the National Center for Missing and Exploited Children (NCMEC) in Alexandria, VA. Under a 2008 federal bill, NCMEC serves as the official clearinghouse for all child pornography and sexual exploitation content reported by electronic communication service providers in the United States. NCMEC is a nonprofit that works as an intermediary between companies and federal law enforcement to review reported images and produce case files for criminal investigation. In 2016, for example, NCMEC reported that they had reviewed over 500,000 image and video files suspected to contain child abuse (NCMEC, 2016). The software Linda was using was PhotoDNA, one of several popular programs for the automated detection of child pornography and the matching of new images across existing databases.

In this paper, I discuss the embodied labor of policing child pornography through the ways in which algorithms and human reviewers like Linda view and sort abuse images. I employ the concept of "apprehension" to suggest that the ways reviewers "see" child pornography is always already oriented toward the capture and arrest of suspected offenders. As I have argued elsewhere (Thakor 2017; Forthcoming), the use of new digital techniques to find child pornography has fundamentally transformed and expanded policing into a distributed network of labor increasingly done by computer scientists and technology companies. Rather than suggest that new software is the

cause of these transformations, I draw attention to the constitutive and mutually defining relation between computing and corporality, or how image detection algorithms need the work of human perception to put their image detection skills to work.

I argue further still that the case study of child pornography detection offers an entry point into examining the algorithmic management of race. I suggest that childhood innocence is coded as whiteness, and whiteness as innocence, in the algorithmic detection of victims and abusers. By taking "detection" as a dynamic practice between human and machine, I make an intervention into critical algorithm studies that have tended to focus solely on the programming of racial bias into software. The algorithmic detection of child pornography hinges, crucially, upon practice and the tacit observation of human reviewers whose instinctual feelings about child protection and offender apprehension become embedded within the reviewing and reporting process as cases escalate for law enforcement.

Digital Forensics

In the past decade many scientists in computer vision have begun directing their research projects beyond facial recognition to full body and scene analysis. Their publications describe the promise of such research for detecting crime scenes and pornographic images. These scenes can be sorted through for skin-color detection, nudity (i.e., the percentage of skin-color pigments visible within a proportion of an image), lighting, scene object recognition, age detection, affect recognition, and kinship matching with known biological relatives. PhotoDNA uses visual mapping and Exchangeable Image File data — a photo's "DNA," according to its designers — from an uploaded image and compares it to other uploaded images. The software uses a form of source identification, allowing "new" images to be compared with images from known databases of child pornography, missing persons, or sex offender registries. The images in the database are assigned a hash value or digital signature. Every file that

has the same data will contain the same hash value. Comparing hash values can provide sufficient evidence that an image is an actual piece of child pornography rather than a fake or an altered photo of an adult. Computer scientists are working on improving the accuracy of such programs (e.g., to rectify distortions or adjust for a tilted face or body through pose correction techniques). Some research teams also work on age progression issues, especially salient for children who may go missing at a young age. A team at the University of Washington has described the craft of interpolating how a person will have aged as "part art, part science and a little intuition" (Kemelmacher-Shlizerman, 2014).

Newer digital forensics projects have their roots in older forms of representational craft in criminal investigations, such as the use of forensic image sketching. The designer of PhotoDNA, Dr. Hany Farid, has been called "the Sherlock Holmes of digital misdeeds" (Adler and McEwen, 2017). Policing institutions have long maintained artists on forensic teams to assist in producing composite images through oral description, to do age progressions on missing persons, or to produce facial reconstructions based on skeletal remains. The composite sketches produced of a victim or a suspected offender are the result of a collaborative and discursive process between investigators, lab experts, witnesses, and family members before becoming manifest through the artist's pencil. The production of a forensic sketch is thus a highly collectivized task, one oriented toward the "apprehension" of a potential victim or offender. Apprehension implies the prospect of arrest or rescue and simultaneously the process of making identities apparent through the procurement of incriminating data. I add to this that in order to know *what* to detect one must become knowledgeable about the object (data, face, person) *sought*.

Forensic investigations have become so increasingly specialized that they fall into the domains of technical personnel. Scholars in science, technology and society (STS) have described this process of professionalization as a "scientification" of the police, with the division of policing into public safety and criminal investigations (Ericson and

Shearing, 1986). While "street cops" devoted to public safety regularly patrol routine areas and occasionally have breakthrough hunches, investigative police are trained to work more acutely on long-term cases and to use forensic technologies, from fingerprinting to DNA testing, to obtain evidence and bolster claims in legal proceedings. However, the way this idealization actually works out is a combination of automated searches with human reviewing. The accumulation and sifting through such images — photos and videos of child abuse and abusers — is explained by social media companies and law enforcement as in some ways a necessary evil to get at the truth. Social media companies hire their own content review teams both onsite and offshore to reach individual and group decisions on the content of specific photos in "borderline" cases they may want to escalate. These viewers rely on a combination of their tacit knowledge and accumulated expertise as they sift through child abuse image content to identify faces and bodies.

Apprehension

I return to the moment of peering over Linda's shoulder. What was interesting to me was the ways in which she would signal various intervals in the software's detection. She would gesture and make side comments on what the program was doing. Linda scrolled down the results display and paused to point out a couple of photos. "See, the child looks a little older here, or at least larger. So this is a problem." It is a problem in the sense that if the child seems to have aged during the course of the reporting, and various photos at different ages are on file, it is more likely the child is still being abused and coerced into posing for such photos. Despite the automated matches, she still continued to pause and point out her own data points in the images. Her remark that "this is a problem" indicated a gut feeling — her tacit knowledge — that some images represented possible other ages and the need for further investigation.

As the anthropologist Charles Goodwin has noted, gestures like

Linda's pointing and verbal cues are embodied communicative modes that structure practice and indicate expertise (Goodwin, 2003). Harry Collins, similarly, points out that interactional expertise helps us understand the dynamic relationship between software and human, and that such expertise is highly embodied and corporeal (Collins, 2016). Image content review is perceptive, interpretive work and also trainable. Once Linda's search query results in several hundred photos, she searches to disqualify any images that are very obviously — to her eye — not a match or do not belong in the set. In doing this disqualification, she helps train the software to produce more accurate results in its next search query, which will be based on the classification elements of the qualifying and disqualifying images.

Such work entails skilled vision, "a capacity to look in a certain way as a result of training" (Grasseni, 2004, p. 41). My understanding of apprehension, then, takes into account how "viewing" becomes a skillful practice through trainings that establish shared ways of "seeing" child-exploitation images.[1] Cristina Grasseni emphasizes that "one never simply looks. One learns *how* to look" (2012, p. 47). Such ways of seeing, distributed and honed across human and computer vision, become manifest as ways of accessing the world and managing it. As police forensic investigations have become increasingly specialized, investigative police are trained to work more acutely on long-term cases and to use forensic technologies, from fingerprinting to DNA testing to image detection software, to obtain evidence and bolster claims in legal proceedings.

I attended one of these training sessions at the Crimes Against Children Conference in Dallas, TX, in August 2014. The conference organizers had coordinated the first annual "Digital Crime Scene Challenge" workshop. In teams of three, participants were instructed to execute a search warrant on a suspected possessor of child pornography in a mock crime scene and produce as much evidence as possible in the fastest time. The game was set up by a special agent from the US Department of Homeland Security and sponsored by a software firm.

Success in the game depended upon a specific visual acuity, recall memory of past cases, efficient collaboration with teammates, and working knowledge of software to analyze confiscated image files. A group of three investigators from Oklahoma won the crime scene challenge, interviewing a person role-playing as a suspect, issuing a search warrant, cataloging evidence in the room, and coordinating next steps for the investigation, including running the confiscated files through a software program to test for any known images of missing or abused children. The use of such training games by police and tech companies focuses the eye to produce shared idioms and ways of seeing, embedding skilled vision into the professional practice of policing.

Viewing Economies

Seeing is also tiered, or classed, so that certain perceptions are afforded higher authority. In these perceptive regimes (cf. Goodwin, 1994), there is an entire viewing economy for how photos of child abuse get "found" and reported: from content moderators in the Philippines or India, to reviewers at tech companies, to NCMEC and federal investigators, and perhaps eventually to international liaisons and INTERPOL.

Indeed, two content moderators working for an agency contracted through Microsoft recently filed a lawsuit against Microsoft, claiming Post Traumatic Stress Disorder (PTSD) from the viewing of violent images in their work (Chen, 2017). Recent work by Sarah T. Roberts and others working in critical digital labor studies have called into question the psychic and emotional costs of monitoring images such as terrorist propaganda, live-streamed executions, and sexual violence against children that get uploaded to sites like Facebook and Twitter every day (Roberts, 2016). This emotional cost is disproportionately borne by people of color working in contingent and precarious contract-labor sites in the Global South and lower-income rural zones in the Global North. Their work is difficult and disposable, with contracts easily dropped and new moderators quickly replacing them.

Many researchers working on image detection algorithms capitalize on this reported trauma to emphasize the possibility of a future where humans may not need to see child abuse images for legal content review, and that it might be entirely mechanized. The marketing of image forensics and classification software often touts its high accuracy and low rate of false positive results; that it will attain a speed and accuracy not possibly by the human eye alone.

Digital Racial Matter

Nonetheless, child pornography detection can never be a fully automated process — despite what digital forensic designers might claim — and will always rely upon human perception and gut since the end goal will always be oriented toward the apprehension of a suspected offender. Designers of algorithms explicitly understand the ways in which *certain* content reviewers, such as Linda, will use and train the software based on their preexisting expertise, and the content reviewers (or their managers, rather), understand the ownership they have over decisions to escalate an investigation.

I thus call the network of various investigators involved in viewing and policing child pornography "algorithmic detectives." Algorithmic detectives symbolize the expansion of policing from traditional law enforcement to a coordinated network of content reviewers, technology specialists, computer programmers, and social media companies who participate in establishing ways of seeing child pornography. Algorithmic detectives do the work of apprehension through training in technical and perceptive skills. There is a constitutive relationship between human vision and computer vision, between computing and corporality. Image detection algorithms *need* the work of human perception to put their detective skills to work. This work must combine human and machine vision with an orientation toward arrest.

Recent work in STS and the social study of algorithms has pointed out, critically, that algorithms have embedded values and biases that can

lead to potential discrimination as well as the erasure of human analysis and input. The accuracy of machine-learning algorithms like image detection is dependent upon the data sets from which algorithms learn. Algorithmic detectives, through their pointing, selecting, and validating of image matches, help train the software to better detect. In the case study of child pornography detection, digital forensics designers use certain training sets to improve their algorithm's ability to detect faces, genders, and races. Historically these training sets have treated the white, male, adult face as a default against which computer vision algorithms are trained (cf. Crawford, 2016). Databases of missing and abused children held by NCMEC in the US disproportionately contain images of white children; non-white children are statistically less likely to be reported as missing or have extensive case files of data. So what happens when certain data and faces are less easily detected, or less accurately matched? For whom might these errors or elisions be of grave consequence?

First, certainly, fewer non-white children might be detected as rapidly as white children, a possibility echoing investigation biases that have persisted since long before the implementation of new software. In the course of my fieldwork, computer vision researchers would often remark that they earnestly wanted to address the "problem" of under-detectability, or even *un*-detectability, of darker skin tones, Black and East Asian features, and younger ages. While they were researching detectability across color tones, light exposures, and eye shape and size, the simpler, more immediate solution they offered was to expand the data set of available faces, an example of what Andrejevic and Gates (2014) have called a "catch all" approach to big data. Indeed, a recent project launched by Facebook in Australia has asked users to submit their nude selfies for Facebook to hash and add to their secure data repository in order to better combat "revenge porn" by ensuring its removal before someone might even post it (Solon, 2017).

But in addition, child victim identification is always conjoined with the imperative to detect suspected offenders. I note that computer vision

researchers' use of the word "problem" might differ from social scientists' perspective on the issue of detectability. As Simone Browne has argued, the inability of computer vision software to distinguish dark skin tones is perilously paired with a hypervisibility tying the history of surveillance to the history of the surveillance of Blackness (Browne, 2014). Are non-white adults, for example, more likely to be marked as suspects because their facial nuances are less accurately detected than white offenders' faces? Steve Anderson urges historians of visual culture to root any analysis of "new" image technologies in the knowledge that visibility has always been tied to logics of colonial control and capture (Anderson, 2017). If certain groups are less accurately detectable, more indistinguishable, and more likely to generate false positives, offender recognition in child pornography cases may be just as capable of producing biases similar to those leading to prejudiced arrests in anti-trafficking cases. As Bernstein (2007) and Woods (2013) have documented, legislation designed to combat sex trafficking has disproportionately resulted in the arrests and prosecutions of Black and Latino men, repeating the carceral logics of other forms of policing and investigation.

Proximity to Innocence

Critically, the expansion of policing practice is made permissible through its use in the case of child protection. The protection of the child hinges upon defining its other (i.e., that which threatens childhood). I argue that algorithmic detectives must thus embody a certain proximity to child victims, to innocence itself, while simultaneously engaging a closeness to culpability, to imagined abusers. Underlying any support of image recognition software is an assumed universal understanding that sexual exploiters of children must be "othered" — located, corralled, and arrested — and are deserving of punishment. It is also understood, then, that investigators who wish to locate and apprehend offenders must first become intimately knowledgeable about sexual abuse — bring it *close*, in a sense — before arresting and enacting punishment. The othering of

potential offenders is a form of extension (cf. Ahmed, 2006), a reaching toward those who must be kept away. Carceral technologies perform this act of extension — apprehending and making proximate violent others — by extending the corpus of policing power in order to seek certain bodies. We might consider them as what Sandy Stone has called a “prosthetic technological self” (Stone, 1991), an extension of a punitive impulse to entrap those who might offend.

The digital forensics software programs I am discussing here are also surveillance technologies. They sort through the massive amounts of text, image, and video data uploaded online every day and work in collaboration with content moderators to keep the Internet secure and "clean." The software itself maintains proximity to innocence. There is a certain pleasure (Magnet and Rodgers 2011) in this mode of surveillance. That is, I am not suggesting that algorithmic detectives gain pleasure from looking at child exploitation images themselves — the pleasure is instead rooted in a moment of recognition toward arrest, the seeing of images for the purpose of policing. Image recognition algorithms, and all the forms of data gathering and surveillance they induce, are justified through the pressing moral obligation and urgency to find and stop child exploitation online. This software has promissory value as a tool for a particular view of *safety and security* — for some, not all.

Recent work in queer feminist social theory has considered how sex offenders have emerged as people who are “despised and disposable” (Horowitz, 2015) or even othered “beyond the pale of humanness” (Borneman, 2015). Child protection surveillance campaigns hinge upon securitizing normative humans through the crafting of *in*-security for inhuman others. Non-police actors who take part in police work strengthen the carceral state by reaffirming the production of *in*security as an exposing practice to shame those labeled criminals or potential offenders. Further still, this exposing practice gets recoded as technical knowledge and moral obligation. Such projects magnify the "stranger danger" panic[2] motivating understandings of sex offenders — the assumption that unknown others are constantly lying in wait to

assault innocent children, an assumption that acquires even more fearful resonances in digital space.

This innocence is becoming quite literally *encoded* as whiteness, with white children's faces better detected, and non-white suspected adult faces more likely to be indistinguishable and less accurately documented. The public effortlessly approves of the capillary forms of surveillance needed to enforce algorithmic detection of child pornography, from Facebook photo albums to increasing partnerships between law enforcement and technology companies. This approval hinges upon the tacit preservation of an imagined docile and innocent symbolic child figure. By bringing child protection *close*, this dispersed network encourages computer scientists, humanitarian professionals, and the general public who view anti-exploitation publicity campaigns to support indiscriminate carceral punishment of sexual exploiters and offenders. In addition, such distributed apprehension expands the policing infrastructure and security thinking of the global north, by working to make race material as digital data, while simultaneously rendering invisible the racialized labor of outsourced content moderators who increasingly bear the emotional toll of human-performed digital janitorial work.

My research suggests that new algorithmic developments fall under a broader phenomenon of expanded digital surveillance that has become amplified by efforts to fight child pornography and trafficking. New digital experts entering into anti-exploitation work have bolstered the child protection agenda traditionally held by law enforcement and anti-trafficking activists, namely, that protection is best achieved through punitive measures. By following the historical trajectory of scientification within law enforcement, as well as current collaborative efforts, we can trace how these actors may embed punitive logic into the investigative process through dual forms of apprehension: image detection software for identifying and locating abuse photos and videos online, and the actual arrests of abusers made possible through partnerships between law enforcement and software companies. Apprehension is justifiable

through an attachment to child innocence. The protection of such innocence relies upon the specter of harm, and entails maintaining a carefully curated proximity to the imagined offenders who might destroy it, in order to locate and arrest these offenders. By examining the particular digital data-sets and legal manifestations of these attachments to protection and punishment, we might begin to see that "the child," in all its virtual and symbolic sense, has perhaps always been white, and the linchpin in the management of whiteness.

Acknowledgements

Thank you to the panel organizers for inviting me to participate in this project, the editors of *Catalyst* for their support, and two anonymous reviewers for their helpful comments.

Notes

[1] As Goodwin explains, "crucial work in many different occupations takes the form of classifying and constructing visual phenomena in ways that help shape the objects of knowledge that are the focus of the work of a profession" (167). Different actors—be they law enforcement investigators, digital forensic startups, social media company reviewer teams, or the outsourced content review workers who are contracted by larger corporations to make the first reviews of potentially abusive images—learn to adapt shared ways of seeing images.

[2] "Some things more than others are encountered as 'to be feared' in the event of proximity, which is exactly how we can understand the anticipatory logic of the discourse of stranger danger" (Ahmed 2004: 40).

References

Adler, Simon and Annie McEwen. RadioLab. (2017). Retrieved from http://www.radiolab.org/story/breaking-news/

Ahmed, Sara. (2004). *The Cultural Politics of Emotion*. Edinburgh, UK: Edinburgh University Press.

Ahmed, Sara. (2006). *Queer Phenomenology: Orientations, Objects, Others*. Durham, NC: Duke University Press.

Bernstein, Elizabeth. (2007). The Sexual Politics of the New Abolitionism. *Differences 18*(3):128-151.

Borneman, John. (2015). *Cruel Attachments: The Ritual Rehab of Child Molesters in Germany*. Chicago, IL: University of Chicago Press.

Browne, Simone. (2015). *Dark Matters: On the Surveillance of Blackness*. Durham, NC: Duke University Press.

Chen, Adrian. (2017). The Human Toll of Protecting the Internet from the Worst of Humanity. *The New Yorker.* Retrieved from https://www.newyorker.com/tech/elements/the-human-toll-of-protecting-the-internet-from-the-worst-of-humanity

Collins, Harry. (2016). Interactional Expertise and Embodiment. In Sandberg, J., Rouleau L., Langley, A., & Tsoukas, H. (Eds.) *Skillful Performance: Enacting Expertise, Competence, and Capabilities in Organizations: Perspectives on Process Organization Studies* (Vol. 7). Oxford, UK: Oxford University Press.

Crawford, Kate. (2016). Artificial intelligence's white guy problem. *The New York Times.* Retrieved from https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

Ericson, R.. & Shearing, C. (1986). The Scientification of Police Work. In Böhme, G. & Stehr N. (Eds.), *The Knowledge Society: The Growing Impact of Scientific Knowledge on Social Relations* (pp. 129-159). Boston, MA: D. Reidel Publishing Company.

Goodwin, Charles. (1994). Professional Vision. *American Anthropologist, 96*(3):606-633.

Goodwin, Charles. (2000). Practices of Seeing Visual Analysis: An Ethnomethodological Approach. In T. Van Leeuwen & C. Jewitt (Eds.), *The Handbook of Visual Analysis* (pp. 157-182). London, UK: Sage Publishing.

Goodwin, Charles. (2003). Pointing as situated practice. In S. Kita

(Ed.), *Pointing: Where Language, Culture, and Cognition Meet* (pp.217–241). Mahwah, NJ: Lawrence Erlbaum.

Grasseni, Christina. (2004). Skilled Visions: An Apprenticeship in Breeding Aesthetics. *Social Anthropology 12*(1):41-55. Pp. 41.

Horowitz, Emily. 2015. *Protecting Our Kids? How Sex Offender Laws Are Failing Us*. Santa Barbara, CA: Praeger.

Kemelmacher-Shlizerman, Ira, et al. (2014). Illumination-Aware Age Progression. Retrieved from https://grail.cs.washington.edu/aging/Aging_CVPR14.pdf

Magnet, S. & Rodgers, T. (2011). Stripping for the State: Whole body imaging technologies and the surveillance of othered bodies. *Feminist Media Studies, 12*(1), 101-119.

National Center for Missing and Exploited Children. (2017). NCMEC Annual Report, 2016. *National Center for Missing and Exploited Children.*

Roberts, S. (2016). Digital refuse: Canadian garbage, commercial content moderation and the global circulation of social media's waste. *Wi: Journal of Mobile Media*, 10(1), pp. 1-18. http://wi.mobilities.ca/digitalrefuse/

Solon, Olivia. (2017). Facebook asks users for nude photos in project to combat revenge porn. *The Guardian. https://www.theguardian.com/technology/2017/nov/07/facebook-revenge-porn-nude-photos*

Stone, Sandy. (1991). Split Subjects, Not Atoms; Or, How I Fell in Love With My Prosthesis. In C. Gray (Ed.), *The Cyborg Handbook* (pp.393-406). New York, NY: Routledge.

Thakor, M. (Forthcoming). Policing Child Exploitation Across Digital Territory. *Science, Technology, and Human Values*. Series on Cybersecurity and Digital Territory: Nation, Identity, and Citizenship.

——— (2017). The Allure of Artifice. In C. Cipolla, K. Gupta, D. Rubin, & A. Willey (Eds.), *Queer Feminist Science Studies: A Reader* (pp. 141-156). Seattle, WA: University of Washington Press..

Woods, T. (2013). Surrogate Selves: Notes on Anti-Trafficking and Anti-

Blackness. *Social Identities, 19*(1), 120-134.

Bio

Mitali Thakor is a postdoctoral fellow in the Sexualities Project at Northwestern, with affiliations in Anthropology and Gender & Sexuality Studies. Her research and teaching cover issues of policing, computing, surveillance, sexual violence, and queer studies of robotics. She is currently working on her book manuscript, *Facing the Child*, an ethnography of artifice, evidence, and the global policing of child pornography. Mitali earned her Ph.D. in 2016 from MIT's Program in History, Anthropology, and Science, Technology, & Society. In Fall 2018, Mitali will join the faculty at Wesleyan University as an Assistant Professor in the Science in Society Program. You can read more about her work at http://www.mitalithakor.com and on Twitter at @mitalithakor.