Empathy in an Age of Deepfakes

Clifford Anderson
Center of Theological Inquiry in Princeton
clifford.anderson@vanderbilt.edu

What is the potential of empathy in helping us see through and beyond deepfakes? Deepfakes are synthetic media that depict individuals acting in falsified circumstances. With growing concern about the propagandistic uses of deepfakes, researchers are actively working on countermeasures to detect synthetic media. This paper examines whether empathy can play a role in differentiating deepfakes from genuine media. After exegeting the phenomenological interpretation of empathy in the works of Edmund Husserl and Edith Stein, the paper explores whether empathy could play a gnoseological role in an interdisciplinary campaign against deepfakes.

What is the potential of empathy for helping us to see through and beyond deepfakes?[1] The present essay continues the analysis of my previous work on synthetic media, focusing on the promise and peril of relying on empathy as a mode of engagement with deepfakes.[2] In my previous essay, I introduced deepfakes, covering the potential benefits and liabilities of this emerging form of synthetic media. In this paper, my goal is to pay due on a conceptual promissory note. Near the conclusion of my previous paper, I contended that 'cultivating empathy' might complement the technological and legal avenues of response to the threat of deepfakes. In what follows, I seek to

[1] This is an incomplete and unpolished draft presented for feedback and discussion at the Theologies of the Digital Conference on April 23-24, 2021.
[2] Anderson 2019.

Zeitschrift für explorative Theology
CZeth_4 (2024) S. 97–117
DOI: https://doi.org/10.17885/heiup.czeth.2023.4.24707
Zuvor veröffentlicht auf PubPub: https://doi.org/10.21428/fb61f6aa.89751048

unpack the relevancy of cultivating empathy for engaging with deepfakes, drawing on the phenomenological tradition but also on writings of golden age, silver age, and contemporary authors of science fiction.

1. Shattering Our Shared Reality

Let's begin by rehearsing the dangers posed by synthetic video. Nina Schick, author of *Deepfakes: The Coming Infocalypse*, considers deepfakes as a new tool in long-running propaganda wars. Harkening back to the Cold War, she cites a Soviet political operative as stating the purpose of a disinformation campaign is to "change the perception of reality."[3] The point of disinformation is to divide the common will of the enemy by undermining their understanding of what is real. In the case of deepfakes, misinformation campaigns seem to have found their perfect weapon. If anyone with a little programming knowhow and some cloud computing credits is able to create realistic videos of their political opponents, who will be in a position to judge the truth or falsity of any video evidence?

The arrival of deepfakes on the media scene served to heighten the already existing challenge of maintaining a shared sense of reality in a digital age. Prior to the Internet age, the sources of knowlege about the past remained relatively limited and fixed. Research libraries, historical societies, and other memory institutions maintained the collections of newspapers, scientific journals, corporate and personal archives, and other documentary records along with the secondary literature required to contextualize and interpret them. Archivists and librarians developed intrinsic standards to arrange, describe, and make these materials accessible to researchers. By contrast, the contemporary media landscape is anything but stable.[4] The effusion of personalized content that algorithms curate and deliver to us renders the establishment of a common historical record increasingly difficult. Common cultural touchstones still exist, however. At the Vanderbilt Television News Archive, for example, video archivists assiduously preserve televised presidential addresses in the conviction that, collectively, these speeches constitute part of our shared cultural experience.

But what would happen if presidential talks were to become personalized? In *Version Control*, Dexter Palmer imagines a scenario where the President of the United States regularly appears on television to share his thoughts about the content of upcoming shows while also dropping in on video screens at restaurants and interrupting

[3] Nina Schick 2020: 55.
[4] Rumsey 2016

video calls to offer political commentary and perspective. Palmer does not explain the mechanism behind these personalized interlocutions, but he hints that they are the equivalent of interactive deepfakes.

> If you had never watched that much television, then you might wonder how it was that the President of the United States had found the time to record a video introduction to every program that appeared on every one of the hundreds of available channels—not just a generic twenty-second speech that gave his imprimatur to the program about to commence, but a short monologue that always seemed to be tailored to the program's subject matter, linking it to some larger political or spiritual meaning. But keen-eyed viewers knew that the President repeated himself: he almost always delivered one of a finite number of canned speeches, perhaps tweaking a word or two in a halfhearted effort at personalization, and anyone who viewed a variety of programs for long enough was bound to see a prologue for a telecast of an English soccer match repurposed a few months later for a stream of a StarCraft I tournament final.[5]

This kind of personalization is not far off. We already receive personalized political emails and robocalls. What Palmer describes is practically feasible already, though not yet culturally acceptable. But we could imagine a future, not so very distant, when political figures deploy data-driven and interactive deepfakes to tailor messages to individual constituents. If such a society were to emerge, what role would archivists play in preserving the past? Perhaps archivists would take to cataloging the recycled stories for future reference and analysis?

We think about deepfakes today primarily as a genre of videos that we encounter on the internet, rendering their subjects' external appearances into puppets that act out the intentions of their creator. As I noted in my earlier contribution on deepfakes, such puppeteering serves both negative and positive ends (as well as mixed ones). But deepfakes are already evolving beyond these simple puppeteering videos to something qualitatively different. I will refer to these prevailing forms of synthetic media as weak in order to contrast them with an emerging form of strong synthetic media.

The emerging wave of strong deepfakes promise to be more dynamic and encompassing. My colleague, Ole Molvig, assistant professor of history at Vanderbilt University, recently created a deepfake of Albert Einstein. His deepfake combined synthetic audio, trained on samples of Einstein's English-language speeches, and synthetic video,

[5] Palmer 2016: 16.

trained on images of Einstein, with sentences generated by GPT-2 (Generative Pre-trained Transformer 2), Open AI's tool for synthetic text generation, which he trained on the corpus of Einstein's English-language publications. Imagine, now, taking this experiment a step further by creating an interactive Einstein that you could connect with on Zoom to ask for assistance with your physics homework. This kind of immersive, interactive deepfake is what I intend by the qualifier "strong."

In the *Reality Game: How the Next Wave of Technology Will Break the Truth*, Samuel Woolley, assistant professor of journalism at the University of Texas at Austin, classifies deepfake videos as a species of "computational propaganda."[6] Computational propaganda takes many forms today, ranging from chatbots operating on social networks to synthetic video on YouTube and beyond. What concerns Woolley is that distinct forms of computational propaganda, if left unchecked, may converge into a multi-sensory virtual reality that to its victims becomes practically indistinguishable from reality itself.

> If we do not take action, we could very well end up with scenarios like this. Digital propoganda is not just biased information, enhanced by automation and bots, that can be read on Facebook group pages or in YouTube comment sections. It is technologically enhanced propoganda that people can see, hear, and feel. In the not so distant future, it could be politically motivated information that is also tasted and smelled.[7]

In his discussion of deepfakes, Woolley discusses a range of "tells" that now make it possible to separate synthetic videos from genuine news media. But he also notes that established practices of investigative reporting form an essential complement to those technological measures. "It is a combination of human and technological strategies," he writes, "that can be brought to bear on this problem."[8]

This contribution focuses on the human side of that equation. How can empathy help us to connect with the other beyond the somatic or technological interface? And how might empathy help us to see through the surface of synthetic media, particularly when it assumes a strong form?

[6] Woolley 2020.
[7] Ibid.: 14.
[8] Ibid.: 126.

2. Phenomenology of Empathy

"Do we ever arrive at an *other* phenomenological I," asks Edmund Husserl in his notes from his lectures during the Winter Semester of 1910-1911 in Göttingen.[9] The philosophy of intersubjectivity posed a challenge to Husserl's phenomenology because, in short, phenomenology seems methodologically to exclude the possibility of including other agents within its ambit. Adopting the standpoint of Descartes's *cogito*, Husserl's phenomenological reduction abstains from empirical investigation to focus on the intentional act that binds subject and object. When I think, I think about something, and the subject and object of thought exist together simultaneously in that relationship of thinking. These so-called intending acts serve as the fundamental data of phenomenological reflection, which explores their modalities while bracketing or putting aside presuppositions or extrapolations about their empirical content. That is, the phenomenological method does not make assumptions about the subject or object of intentional acts that transcend what those acts themselves reveal. The bracketing of the so-called naturalistic attitude bolsters the claim of phenomenological investigations to irrefragability. Just as with the *cogito*, I cannot doubt that my intentional acts of experience, the so-called *cogitationes*. But therein lies the rub. When I encounter another person, I must bracket the existence of that person as a conscious agent to explore what is given to me solely within the confines of the phenomenological reduction. Husserl asks whether the phenomenological reduction requires him to see only an animate body where he naturalistically experienced a fellow human being?

Briefly stated, the phenomenological question about empathy asks how to understand our primordial encounter with another when we do not have unmediated access to the "I" of the other. That is, it seeks to understand the possibility of a middle way between two alternatives that it rules out. The first alternative is a kind of solipcism, whereby I do not encounter anyone else directly, but only material forms in motion. I may treat these living objects as creatures and may posit that they have consciousness in a form similar to my own, but I am never directly aware of their "I"s. While Husserl could speak of phenomenology as a kind of methodological solipcism, he was at pains to deny that phenomenology ineluctably led to ontological solipsism. This was the argument, in broad strokes, of his *Cartesian Meditations*, which begins from the standpoint of methodological solipicism and concludes with a magisterial consideraton of intersubjectivity. On the other hand, Husserl and his followers also rejected any mystical or psychical solution that posited direct interaction between "I"s. If I

[9] Husserl 2006: 82.

perceive another person as an "I," it is not because I have direct access in any form to that person's consciousness. My perception of the other as other is primordial but always mediated. This is the riddle that phenomenologists aspired to solve.

Husserl's response to this quandary is to draw on the experience of empathy. "In empathy," he explained, "the empathizing I experiences the inner life or, to be more precise, the consciousness of the other I."[10] In everyday social life, empathy connects us with our fellow human beings. The "I" "sees the other"I's" not in the sense that it sees itself or experientially finds itself," clarified Husserl. "Rather it posits the immanence of 'empathy'; hence other lived experiences and other character dispositions are 'found' too; but they are given or had in the sense of one's own."[11] Empathy designates our ability to experience the other as another I rather than as an animate body. Husserl emphasizes that empathy is not about mirroring the activities of the other– for instance, feeling angry when the other radiates anger. As he noted, "For when I feel empathy with your anger, I am myself not angry, not at all."[12] The relationship is more primordial; in empathy, we experience the "I" along with the physical body. Husserl argues that the experience of empathy, like any intentional act, survives the phenomenological reduction. The reduction allows us to explore the intentional act of empathy *qua* act, relating not to this or that particular individual, but to the experience, indubitable in its own right, of perceiving a fellow I in, with, and through a physical body.

The description of Husserl's nascent phenomenology of intersubjectivity must suffice for the present purpose. A detailed explanation would have to trace the development of his ideas about empathy from his earliest work to his *Cartesian Meditations* and beyond. In his *Nachlass*, Husserl left behind manuscripts that Iso Kern painstakingly reconstructed and published as *Husserliana* XIII–XV[13] Husserl was evidently also dissatisfied with the crabbed exposition in his lectures from 1910-1911, rewriting them with more precise philosophical terms (that again raise new questions) in what is now Appendix XII of the volume.[14] In the course of his meditations on the philosophy of intersubjectivity, Husserl benefited from conversations with his doctoral student, Edith Stein. Edith Stein was born to a Jewish family in Breslau, Silesia (now Wrocław, Poland) and had initially studied psychology. In 1913, she arrived in Göttingen to at-

[10] Ibid.
[11] Ibid.: 5.
[12] Ibid.: 83.
[13] McCormick 1976: 167–89.
[14] Husserl 2006: 157–64.

tend Husserl's lectures on phenomenlogy, hoping to discover a better undergirding for psychology as a science.[15] While her studies were interrupted by the oubreak of the war in 1914, by 1915 she had returned to study with Husserl, by now a professor in Freiburg, and eventually completed her dissertation under his guidance in 1916. Her thesis, *On the Problem of Empathy*, set out the fundamental problem of intersubjectivity from a phenomenological perspective and articulated a more thorough exploration of the critical concept of empathy than Husserl had theretofore provided.

In a remarkable passage early in her thesis, Stein noted that understanding interpersonal empathy opens up a window to grasping other forms of empathy, including divine empathy.

> This experience which an "I" as such has of another "I" as such looks like this. This is how human beings comprehend the psychic life of their fellows. Also as believers they comprehend the love, the anger, and the percepts of their God in this way; and God can comprehend people's lives in no other way.[16]

Husserl, by contrast, had opined in 1910-1911 that God had no need of empathy because God had direct insight into the consciousness of all conscious agents, a theological thesis that he termed "divine all-consciousness."[17] Whether empathy connects human beings to other creatures and their Creator remains a central question.

3. The Shifting Semantics of Empathy

Exploring the concept of empathy requires us to attend to its philological evolution. The term 'empathy' is a nineteenth century neologism that, for most of its existence, stood in want of clear definition. As Susan Lanzoni chronicles in *Empathy: A History*, the semantics of the term shifted as researchers from different fields, ranging from aesthetics to psychology to neuroscience, layed claim to the word and attempted to pin down its definition.[18] Most straightforwardly, the English word "empathy" originated as a translation of the German word, 'Einfühlung.' As Lanzoni demonstrates, the term 'empathy' shifted gradually from meaning the projection of oneself into another, whether object or person, to a receptive meaning. "Rather than an expansion

[15] Borden 2004: 4.
[16] Stein 1989: 11.
[17] Husserl 2006: 177–78.
[18] Lanzoni 2018.

of the self into a form or shape, empathy came to mean the very opposite," she explains, namely, "the reining in of the self's expressiveness to grasp another's emotion in service to a therapeutic goal or moral imperative."[19]

Edmund Husserl and Edith Stein's phenomenological explorations of empathy may also be situated in the history of this gradual semantic transformation. From their references to the psychologist Theodor Lipps (1851–1914), we ascertain how they took as their point of departure the aesthetic tradition of empathy while also pushing back against its narrow philosophical frame. According to Montag, et. al., Lipps developed his understanding of 'Einfühlung' from David Hume's concept of sympathy in *A Treatise of Human Nature*.[20] Lipps experimented with methods to demonstrate how the "I" projects itself into objects (for example, seeing movement in certain forms of optical illusions when the lines remains stationary) as well as people (for instance, experiencing fear when watching a circus performer walking a tightrope). As Lanzoni describes, this theory of empathy, which posited that spectators of artwork come to appreciate those works of art by projecting their subjectivity into them, formed the basis of a dominant theory of aesthetics in the early twentieth century.[21]

A counterveiling understanding of empathy began to emerge in psychological circles during that era. "The psychotherapeutic rendering of empathy traded self-projection for its opposite," writes Lanzoni. "One now had to bracket the self's feelings and judgments in order to more fully occupy the position of another."[22] This perspective on empathy became familiar in the form of Rogerian or "person-centered therapy," in which the therapist aspires to empathize with their clients' self-understanding to help clients grapple with and overcome their psychological quandaries.

Different senses of empathy continue to coexist. "Truth be told," admits Lanzoni, "there is little agreement today among psychologists, neuroscientists, and philosophers on empathy's contours."[23] A phenomenological theory combines aspects of both the projective and receptive side of empathy. In exploring the relation of empathy to deepfakes, we may also find that both dimensions are necessary. If we project ourselves into the other, we seek to humanize the technological object. But when that projection fails to encounter any genuine I beyond the somatic appearance of the self, the empathizer may recoil and revoke their extension of empathy.

[19] Ibid.: 14.
[20] Montag / Gallinat / Heinz 2008: 1261.
[21] Montag / Gallinat / Heinz 2008: Chapter 3.
[22] Lanzoni 2018: 125.
[23] Ibid.: 252.

4. The Empathy Snatchers

The most famous work of deepfake science fiction is undoubtedly Jack Finney's 1955 novel, *The Body Snatchers*, now better known as *Invasion of the Body Snatchers* after its multiple film adaptations. The novel portrays the arrival of interstellar parasites in the fictional town of Mill Valley, set in Marin County, California of the 1950s. The protagonist, Miles Bennell, is a local physician. At the beginning of the novel, he receives an after hours visit from Becky Driscoll, who reports that a close friend has become convinced that her uncle Ira is not actually her uncle.

> "Miles, she's got herself thinking that he *isn't* her uncle." "How do you mean?" I took a sip from my glass. "That they aren't really related?" "No, no." She shook her head impatiently. "I mean she thinks he's"–one shoulder lifted in a puzzled shrug–"an imposter, or something. Someone who only *looks* like Ira, that's all. Miles, I'm worried sick!"[24]

The characters in the novel assume, at first, that the town is experiencing a kind of mass psychosis, a frightening but transient delusion. What becomes evident as the action continues is that an alien lifeform is spreading through the town, planting pods in people's basements and closets, which eventually replicate and destroy their human hosts. The pod people look, act, and speak identically to the originals. They share the same memories, making it easy for them to blend in. But while they can mimic emotion, they do not themselves have any emotions. The lack of affect is the only "tell."

> "There was only one way Wilma Lentz knew Ira wasn't Ira. Just one way to tell, because it was the only difference. There was no emotion, not really, not strong and human, but only the memory and pretense of it, in the thing that looked, talked, and acted like Ira in every other way."[25]

Given their emotional vacuity, the pod people lack the ability to empathize with human beings. As with contemporary deepfakes, the eyes prove the most difficult to emulate and, on the flip side, the most revealing of the hollowness within the replicants. Finney focuses on the alterity of the gaze in his description of the encounter of Miles and Becky with the town librarian.

---

[24] Finney 2010: 11.
[25] Ibid.: 184.

> For a moment she still stood, glancing helplessly from me to Becky in utter bewilderment; then suddenly she dropped the pretense. Gray-haired Miss Weygand, who twenty years ago had loaned me the first copy of *Huckleberry Finn* I ever read, looked at me, her face going wooden and blank, with an utterly cold and pitiless alieness. There was nothing there now, in that gaze, nothing in common with me; a fish in the sea had more kinship with me than this staring thing before me. Then she spoke. *I know you,* I'd said, and she replied, and her voice was infinitely remote and uncaring. "Do you?" she said, then turned on her heel and walked away.[26]

While the replicants lack empathy for their human hosts, Miles and Becky continue to feel empathy for their lost friends and relations, finding it difficult to strike and kill the pod people who impersonate them so nearly.

In the final section of the novel, the clones discover and trap Miles and Becky in his medical office off the town square. Finney uses this scene to explore the clones' perspective. What makes the clones frightening is not their malevolence, but their utter lack of caring. Finney expertly turns this lack of empathy back on his readers.

> "You look shocked, actually sick, and yet what has the human race done except spread over this planet till it swarms the globe several billion strong? What have you done with this very continent but expand till you fill it? And where are the buffalo who roamed this land before you? Gone. Where is the passenger pigeon, which literally darkened the skies of America in flocks of billions? The last one died in a Philadelphia zoo in 1913. Doctor, the function of life is to live if it can, and no other motive can ever be allowed to interfere with that. There is no malice involved; did you hate the buffalo? We must continue because we must; can't you understand that?" He smiled at me pleasantly. "It's the nature of the beast."[27]

The passage obviously points back to the reader, questioning us about our lack of empathy for other species. Are we simply beasts in the end, with empathy serving as nothing more than an evolutionary adaptation benefiting the survival of the human race? If so, might the future course of evolution favor empathyless androids who have transcended human emotional limitations? With that dismal thought in mind, we turn to a classic of the silver age of science fiction.

---

[26] Ibid.: 129.
[27] Ibid.: 187.

5. Do Androids Empathize with Electric Sheep?

The novelist Philip K. Dick (1928–1982) gave the 'imitation game' a new and deadly twist in *Do Androids Dream of Electric Sheep*? Dick depicted a future in which a commercial firm produces android servants for space colonists, but cannot under penalty of law import them to earth. As the models develop, these androids become virtually indistinguishable from human beings. Rebelling against their sidereal enslavement, a few androids from the latest Nexus-6 line manage to escape their bonds and flee to earth. Rick Deckard, a bounty hunter, must hunt them down and "retire" them.

At the beginning of the novel, Deckard muses that intelligence no longer serves to distinguish the latest androids from humans. These androids have long since passed the Turing Test. "Well, no intelligence test would trap such an andy." But Deckard can make use of a new heuristic, the so-called "Voigt-Kampff Empathy Test."

> [Deckard] had wondered, as had most people at one time or another, precisely why an android bounced helplessly about when confronted by an empathy-measuring test. Empathy, evidently, existed only with the human community, whereas intelligence to some degree could be found throughout every phylum and order including arachnida.[28]

Contemporary researchers have also proposed testing machines for empathy. In "An Empathy Imitation Game: Empathy Turing Test for Care- and Chat-bots," Jeremy Howick, Jessica Morley, and Luciano Floridi argue that machines must show empathy to operate effectively in environments like clinical settings. A patient would presumably resent being informed of a fatal condition by a robot that ended the announcement with a cheery, 'Have a nice day!' "We propose to move this debate from the abstract to the concrete," they write. "Taking our inspiration from the Turing Test for human thinking..., we propose to replace 'can artificial carers be empathic?' with 'can a human user distinguish between the empathy showed by an artificial carer and that showed by a human practitioner?' "[29] Selecting a standard instrument for measuring patients' perceptions of caregivers' empathy, the authors contend that, given suitable modifications, the tool could also assess whether artificial caregivers exude empathy toward their subjects of care. By studying whether artificial agents are able to achieve levels of empathy equivalent to human caregivers, they express hope, while

[28] Dick 1996: 29.
[29] Howick / Morley / Floridi 2021.

allowing that ethical concerns about deceptive empathy exist, that "philosophical debates about the extent to which artificial carers can be empathic [may be] sidestepped in favour of rigorous Turing-type tests that compare perceived empathy of a care or chatbot with perceived empathy of a human practitioner."[30]

The difference between the fictional Voigt-Kampff test and the real world instrument for measuring empathy is the directionality, that is, who is primarily being assessed. In the Voigt-Kampf test, the agent seeks to suss out androids by assessing the genuineness of their surface empathy. In the Howick-Morley-Floridi proposal, by contrast, the administrator would measure the extent to which a patient has been taken in by artificial expression of empathy. Their assumption seems to be that, at least in certain clinical circumstances, the appearance of empathy suffices.

Do Androids Dream of Electric Sheep? is at heart a reflection on empathy. Androids can fake empathy, but they are not genuinely empathatic. They do not care about human beings. The androids regard empathy as a human weakness. Human beings, by contrast, appear driven to extend empathy beyond their kin. In an echo of Finney's description of the alien librarian in *The Body Snatchers*, Dick describes the moment a human unwittingly discerns that his new neighbor is different than her appearance.

> Now that her initial fear had diminished, something else had begun to emerge from her. Something more strange. And, he thought, deplorable. A coldness. Like, he thought, a breath from the vacuum between inhabited worlds, in fact from nowhere....[31]

The human characters in the novel adhere, with greater or lesser devotion, to a religion of empathy called "Mercerism." The religion centers on empathetic identification with an individual, perhaps historical, perhaps archetypal, named Wilbur Mercer, whom anonymous "killers" have cast to the depths of a pit and who seeks, in the face of their taunts and stones, to climb out again, restoring other dead creatures to life as well. Dick offers tantalizing details about Mercerism and, indeed, the shadowy figure of Wilbur Mercer intervenes crucially in the narrative. The androids despise Mercerism as it epitomizes their lack of humanity. The androids believe that by revealing Mercerism as founded on a set of deepfake videos, they can likewise expose empathy itself as fraudulent. " 'Mercerism is a swindle,' " the *de facto* leader of the

[30] Ibid.
[31] Dick 1996: 63.

band of fugitives declares, "The whole experience of empathy is a swindle."[32] The inability of historical-critical evidence to shake the foundations of Mercerism frustrates the androids, as the failures of analogous attempts stymies critics of religion today. Is Dick hinting that the extension of empathy inevitably extends to others, binding humanity in mystical unity?

Dick does not explain the prohibition of androids on earth. Without reading too much of our theme into his narrative, he suggests that civil authorities imposed the ban to avoid the consequences of strong deepfakes. The title of the novel points to the paradox of empathy, the attempt to connect emotionally with unfeeling machines. In the earth of 1992, animals have nearly become extinct. While a privileged few can afford to own an animal, the majority must make do with artificial surrogates. The remaining middle class on earth content themselves with caring for mechanical animals. Deckard once owned a genuine sheep but, when the sheep died, he purchased an artificial surrogate to take its place. While he fools his neighbor, he cannot deceive himself and has come to hate the robotic animal. " 'The tyranny of an object,' he thought. 'It doesn't know I exist. Like the androids, it had no ability to appreciate the existence of another.' "[33] Is there any way to overcome this kind of empathy deficit? If human beings recoil at this lack of mutual appreciation, is the problem insoluble or might there be a different way of compensating for this lack?

6. Compensating for Empathy Deficits

If the perception of emotional hollowness at the core of synthetic media serves as the most fundamental tell that something is a strong deepfake, how should we respond when we detect such a lack? Is there any way to overcome that deficit?

Simon Baron-Cohen, professor of developmental psychopathology at the University of Cambridge, explores how the absence of empathy underlies cruelty and other asocial actions in *The Science of Evil: On Empathy and the Origins of Cruelty*.[34] Baron-Cohen, relying in part on Martin Buber, argues that failure of empathy reduces interpersonal encounter between subjects to the relation between the "I" and an object. As he sees it, the reduction of the other to an object serves as a necessary, if not sufficient, condition for treating the other cruelly. Baron-Cohen does not cite Immanuel Kant, but his reflections echo Kant's second formulation of the categorical imperative,

[32] Ibid.: 210.
[33] Ibid.: 42.
[34] Baron-Cohen 2011.

namely, "So act that you use humanity, in your own person as well as in the person of any other, always at the same time as an end, never merely as a means."[35] But what happens when another agent cannot recognize the emotional state of others? In that case, are they doomed to violate the laws of morality, causing harm to those around them? Not necesarily, argues Baron-Cohen. In certain circumstances, there are ways to overcome empathy deficits, at least of a particular kind.

Baron-Cohen contends that human beings "*all lie somewhere on an empathy spectrum* (from high to low)."[36] He explores the psychopathology of what he terms "empathy erosion," namely, the diminishment of the ability to understand the perspective of other people, that is, to see and sympathize the world from their point of view. Baron-Cohen makes a distinction between two fundamental types: 'Zero Degrees of Empathy Negative' and 'Zero Degrees of Empathy Positive.' The first is, as the name indicates, always negative and, frequently eventuates in harmful and destructive actions. The second form, Baron-Cohen argues, gives beneficial expression to this deficit through compensatory actions. In particular, this form emerges for those who suffer from a lack of "cognitive empathy," that is the ability to understand why a person is feeling the way there are, but who feel "affective" empathy, namely, a sense of care for another person's emotional state.[37] According to Baron-Cohen, such people may make up for that lack by 'systematizing,' which he defines as "the ability to analyze changing patters, to figure out how things work."[38] When applied to the field of ethics, a systematizer prefers to operate with universalizable moral principles rather than contextual ethical guidelines. For Baron-Cohen, this explains why people with Apserger Syndrom (whom, he contends, have deficits in cognitive empathy but maintain affective empathy) "are often the first to leap to the defense of someone who is being treated unfairly because it violates the moral system they have constructed through brute logic alone."[39]

As noted in the reviews of his publication, a controversial aspect of Baron-Cohen's work is his claim that autism, at root, stems from a failure of the empathic circuit to develop normally in the brain. As a critic has cautioned, "a critical autism studies has the potential to alert us to the ideological functions that can be performed when we try to define autism and its relation to notions put forth as 'fundamental human charac-

[35] Kant 2011: 87.
[36] Baron-Cohen 2011: 17.
[37] Ibid.: 109.
[38] Ibid.: citations omitted.
[39] Ibid.: 128.

teristics.' "[40] What matters for our purposes, however, is Baron-Cohen's notion that moral systematizing can overcome deficits in cognitive empathy. If an artificial agent could substitute a set of ethical principles in place of cognitive empathy, would that artificial agent be able to function socially among human beings? The answer seems to be a qualified 'yes.' Indeed, researchers may aim to endow machines with cognitive empathy, that is, the capacity to recognize and respond appropriately to human sentiments while regarding affective empathy as either unwanted or beyond the technological pale.[41]

7. Empathy in a Technological Age

Sherry Turkle, Professor of the Social Studies of Science and Technology in the Program in Science, Technology, and Society at MIT, is our foremost ethnologist of the digital age. A primary focus of her work is the deleterious effects of technology on our capacity for empathy with fellow human beings. In *The Empathy Diaries*, Turkle reflects on how her research interests developed from formative childhood experiences.[42] In particular, her biological and adoptive fathers' failure (or perhaps inability) to consider the world from her perspective, to take her feelings and aspirations into account, led Turkle to the exploration of empathy, and its absence, in her scholarship. In her biography, she shares the personal and academic itinerary that carried her from Brooklyn, to Radcliffe, to studying the psychoanalysis of Jacques Lacan in Paris, to her professorship at MIT. From the campus of MIT, she has defended the role of interpersonal empathy in an increasingly technological society.

"We must confront the downside of living with the robots of our science fiction dreams," she writes. "Do we really want to feel empathy for machines that feel nothing for us?"[43] Turkle calls the effort to create machines with the pretension of empathy "the original sin of artificial intelligence."[44] Is our willingness as humans to extend empathy to non-empathetic agents a flaw or feature of our emotional makeup? Or is the answer perhaps that it is both at once, and that calibrating our response to the situation proves more challenging than opting either for callousness toward or naive comity with the machines in our lifeworld.

[40] McDonagh 2013: 44.
[41] Stephan 2015.
[42] Turkle 2021.
[43] Ibid.: 345.
[44] Ibid.

Looking back at *Do Androids Dream of Electric Sheep*, Deckard worries, on the one hand, that fellow bounty hunter Phil Resh relishes the experience of "retiring" androids, exhibiting zero empathy toward them in a manner that Deckard regards as vaguely psychopathic. On the other, he senses that his own developing sense of empathy toward androids is liability. "Emapthy toward an artificial construct? he asked himself? Something that only pretends to be alive?"[45] In fact, Rachel Rosen, an android manufactured and then employed by the Rosen Corporation, attempts to neutralize Deckard as a bounty hunter by enlarging his sense of empathy for her and her kind. Still, there is something human about empathizing with the unempathetic. Philip K. Dick once remarked, "to me, the ... replicants are deplorable. They are cruel, they are cold, they are heartless, they have no empathy, which is how the Voigt-Kampf test catches them out, don't care what happens to other creatures."[46] But he then went on to observe that "the theme of my book is that Deckard is dehumanized through tracking down the androids."[47] In phenomenological terms, the androids' near perfect mimicry of the somatic behaviors of the human ineluctably engages our analogical sense of empathy. We cannot help *but* seek for the corresponding I of the other. As Deckard realizes, shutting off the effort at empathy is as dangerous as failing to perceive that the android, at its core, lacks the capacity for empathic response.

8. Empathy: Human and Divine

In his *Aids to Reflection*, Samuel Taylor Coleridge meditated on the meaning of James 1:25 "But those who look into the perfect law, the law of liberty, and persevere, being not hearers who forget but doers who act—they will be blessed in their doing" (NRSV). Considering the metaphor of "looking" into the law in Aphorism XXIII, Coleridge noted, "*Quantum sumus, scimus*. That which we find within ourselves, which is more than ourselves, and yet the ground of whatever is good and permanent therein, is the substance and life of all other knowledge."[48] Ralph Waldo Emerson subsequently reversed the phrase to read, *Quantum scimus sumus*, that is, "What we know, we are."[49] If we take empathy as a form of knowing, that is, as an intentional activity that engages us with the world, the more we empathize the more empathetic we become. And, of course, the less empathy features as a primordial form of engagement,

[45] Dick 1996: 141.
[46] Sammon 1981: 27.
[47] Ibid.
[48] Coleridge 2017: 30.
[49] Emerson 1978: 118.

the less caring we become toward others. The danger of seeking to empathize with machines is potentially that we will experience our own empathy eroding. Or, to follow Turkle's more subtle reasoning, "These days, our technology treats us as though we were objects and we get in the habit of objectifying one another as bits of data, profiles viewed. But only shared vulnerability and human empathy allow us to truly understand one another."[50] As we connect more and more with humans through machine interfaces, is technology exercising a corrosive effect on our ability to express and experience empathy?

The potential of technology to foster empathy erosion carries us to the final work of science fiction I wish to consider, namely, Kazuro Ishiguro's *Klara and the Sun*.[51] The primary theme of this novel is also empathy, but the roles have been reversed. Klara is an artificial friend, an android created for the vocation of serving as a companion to children whose parents, presumably, lack the time or the inclination to care for them themselves. Klara belongs to an older class of artificial friend, the fourth generation of the B2 line, which lacks the somatic and cognitive upgrades of the newer B3 model, but which remains unsurpassed in its ability to empathize. Klara seeks to understand the interpersonal world around her, first observing bypassers from a shop window and then learning from the interactions of the stressed family she winds up living with. While Klara comes to greater awareness of limitations and motivations of the human beings in her sphere, the trajectory of the human agents runs the opposite way. The biotechnical artifices they use to boost the cognitive capacity of their children seems to render them steadily less empathetic about others.

The juxtaposition of the empathetic android and the unempathetic humans drives the drama, but a secondary theme of the novel concerns the role of the divine. While the human characters have lost any sense of religiosity, Klara personifies the sun as a benevolent deity and, at crucial junctures in the narrative, petitions the sun to intercede on behalf of others. The sense of strength gained from her faith in the loving-kindness of the sun inspires the human beings around her with hope, even as they instictively disregard the source of her confidence as "well, [Artificial Friend] supersition"[52] The empathy that Klara feels for her young ward leads her to appeal to the empathy of the sun for situation of the child. While Ishiguro leaves the efficacy of the android's religious convictions an open question, he underscores through the contrast between

[50] Turkle 2021: XIX.
[51] Ishiguro 2021.
[52] Ibid.: 287.

Klara and the humans she interacts with how intrincally connected, and connecting, are religious devotion and empathy for others.

9. Two Concluding Codas on Empathy

Two more quick reflections in closing. In her recent young adult novel, *Deepfake*, Sarah Darer Littman examines the potential of deepfakes to disrupt contemporary students' lives. She imagines two students, Dara Simmons and Will Halpern, competing to become valedictorian of their high school class while secretly also dating on the side. After the couple both receive early acceptances to highly-selective colleges, a video posted to an anonymous gossip site shatters their idyll. In the video, Dara offhandedly claims that Will cheated on his SAT to gain admission to Stanford. In the student center of the high school, Will confronts a bewildered Dana, who contends she never said any such thing.

> "I know it looks like I did," I say, breathless desperation making my voice unnaturally high-pitched. "I don't understand how, because I swear that *I never said those things*. /"Yeah? So what's your brilliant explanation for the video showing you doing exactly that?" / ... / "I don't have one," I'm forced to admit. "I don't know where that video came from." I try to put my hand on his arm, my eyes pleading with him, but he flinches away from me. "Will, please...you *know* me. You've got to believe that I am telling you. I would *never* do this to you. /"Except you did," he says. His gray eyes are glacial. "And now I don't know what to believe about you anymore. We're done. Finished. Over."[53]

The plot of the novel unfolds like a detective story. Dara must figure out the origin of the video to clear her name and to salvage Will's college acceptance. Without giving away details of the plot, the video turns out to be a deepfake. Dara is able to identify several convincing "tells" that, when scrutinized, eventually lead her and Will to identify its creator. At a deeper level, the theme of the novel is about trusting despite appearances. In this case, the young love between the couple does not survive Will's disbelief that the video could be anything other than genuine. The deepfake shut down his ability to empathize with Dara, deafening him to her pleas that he *knows* her.

A second case caused a stir in April 2021. VICE posted an interview between Eliza Mcphail, an intern, and Matt Loughrey, a digital artist. The article detailed

[53] Littman 2020: 27.

Loughrey's colorization of photographs of victims of genocide who perished at Security Prison 21 in Phnom Penh from 1976 to 1979. As Loughrey described his motivation for colorizing the images, "It's somewhere between curiosity and empathy."[54] The interviewer remarked on the eerie smiles on the visages of some prisoners. How could they be grinning in the face of their imminent executions? Loughrey offered a pop psychological explanation, but the actual explanation seems to have been more straightforward: Loughrey allegedly retouched the faces of the prisoners' synthetically. The Tuol Sleng Genocide Museum, located on the site of the former Security Prison 21, issued a statement requesting "researchers, artists and the public not to manipulate any historical source to respect the victims."[55] VICE retracted the article, apologizing and promising an editorial investigation.[56] The incident, however, underscores the dangerous appeal of empathy. By drawing us into the story, our empathic nature seeks to understand and engage with the disturbing emotions manifested in the photographs. But, at the same time, we have to attend to our inner sense of dissonance when we cannot imagine ourselves feeling that way when putting ourselves in the place of the other. Such a failure of empathy may arise because someone is manipulating us.

Believing in spite of appearances. Disbelieving despite the evidence. Seeking for signs or "tells" contradictring what otherwise appears genuine. Keeping faith in the other, but not falling prey to false messiahs. Are these traits of empathy in an age of deepfakes? Or characteristics of witnessing to the coming Kingdom of God in a fallen world?

Bibliography

Anderson, Clifford. 2019. "A New Hermeneutics of Suspicion? The Challenge of Deepfakes to Theological Epistemology". *CZeth 3*. https://doi.org/10.21428/fb61f6aa.771d30b7.

Baron-Cohen, Simon. 2011. *The Science of Evil: On Empathy and the Origins of Cruelty*. New York: Basic Books.

Borden, Sarah. 2004. *Edith Stein* (Outstanding Christian Thinkers). London: Continuum.

[54] Eliza Mcphail 2021.
[55] BBC 2021.
[56] Vice Staff 2021.

BBC. 2021. "Cambodia Criticises Edited Photos of Khmer Rouge Victims." *BBC News*. Accessed November 7, 2022. https://www.bbc.com/news/world-asia-56707984.

Coleridge, Samuel Taylor. 2017. *Aids to Reflection* (Edited by John B. Beer. Vol. 9. The Collected Works of Samuel Taylor Coleridge). Princeton: Princeton University Press.

Dick, Philip K. 1996. *Do Androids Dream of Electric Sheep?* Toronto: Del Rey.

Emerson, Ralph Waldo. *Journals and Miscellaneous Notebooks of Ralph Waldo Emerson, 1854–1861* (Edited by Susan Sutton Smith and Harrison Hayford. Journals & Miscellaneous Notebooks of Ralph Waldo Emerson 14). Cambridge: Belknap Press, 1978.

Finney, Jack. 2010. *Body Snatchers*. London: Gollancz.

Howick, Jeremy, Jessica Morley, and Luciano Floridi. "An Empathy Imitation Game: Empathy Turing Test for Care- and Chat-Bots." *Minds and Machines*, February 2021. https://doi.org/10.1007/s11023-021-09555-w.

Husserl, Edmund. *The Basic Problems of Phenomenology: From the Lectures, Winter Semester, 1910-1911* (Translated by Ingo Farin and J. G. Hart). Dordrecht: Springer, 2006.

Ishiguro, Kazuo. 2021. *Klara and the Sun*. New York: Knopf.

Kant, Immanuel. 2011. *Groundwork of the Metaphysics of Morals: A German-English Edition* (Edited by Jens Timmerman. Translated by Mary Gregor). Cambridge: Cambridge University Press.

Lanzoni, Susan. 2018. *Empathy: A History*. New Haven: Yale University Press.

Littman, Sarah Darer. 2020. *Deepfake*. New York: Scholastic Press.

McCormick, Peter. 1976. "Husserl and the Intersubjectivity Materials". *Research in Phenomenology* 6: 167–89.

McDonagh, Patrick. 2013. "Autism in an Age of Empathy: A Cautionary Critique." In *Worlds of Autism: Across the Spectrum of Neurological Difference*, edited by Joyce Davidson and Michael Orsini, 31–52. Minneapolis: University Of Minnesota Press.

Mcphail, Eliza. 2021. "These People Were Arrested by the Khmer Rouge and Never Seen Again." *Vice*, April 2021.

Montag, Christiane, Gallinat, Jürgen and Heinz, Andreas . 2008. "Theodor Lipps and the Concept of Empathy: 1851-1914." *The American Journal of Psychiatry* 165 (10): 1261. https://doi.org/10.1176/appi.ajp.2008.07081283.

Palmer, Dexter. 2016. *Version Control*. New York: Vintage.

Rumsey, Abby Smith. 2016. *When We Are No More*. New York: Bloomsbury Press.

Sammon, Paul. 1981."The Making of Blade Runner." *Cinefantanstique* 12 (5/6): 20–46.

Schick, Nina. 2020. *Deepfakes: The Coming Infocalypse*. New York: Twelve.

Stein, Edith. 1989. *On the Problem of Empathy* (Translated by Waltraut Stein. 3rd Revised edition. The Collected Works of Edith Stein). Washington, D.C: ICS Publications.

Stephan, Achim. 2015. "Empathy for Artificial Agents." *International Journal of Social Robotics* 7 (1): 111–16. https://doi.org/10.1007/s12369-014-0260-0.

Turkle, Sherry. 2021. *The Empathy Diaries: A Memoir*. New York: Penguin Press.

Vice Staff. 2021. "Editorial Statement Regarding Photographs of Khmer Rouge Victims." *VICE*, April 2021.

Woolley, Samuel. 2020. *The Reality Game: How the Next Wave of Technology Will Break the Truth*. New York: PublicAffairs.