This is a section of doi:10.7551/mitpress/14471.001.0001

The Perception Machine

Our Photographic Future between the Eye and AI

By: Joanna Zylinska

Citation:

The Perception Machine: Our Photographic Future between the Eye and AI
By: Joanna Zylinska
DOI: 10.7551/mitpress/14471.001.0001
ISBN (electronic): 9780262376631
Publisher: The MIT Press
Published: 2023

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open

The MIT Press

4 From Machine Vision to a Nontrivial Perception Machine

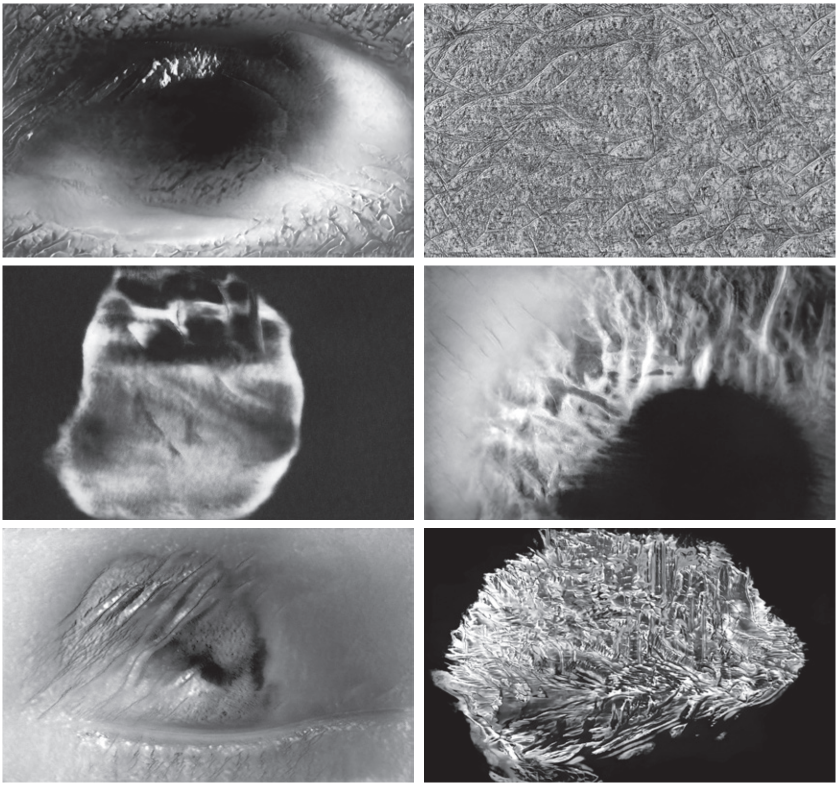

Figure 4.1
Stills from Joanna Zylinska, *Neuromatic*, 2020.

Neuromatic

The images that open this chapter (figure 4.1) come from a video work of mine called *Neuromatic*, which frames the key issues of the chapter. To make the video, I reanimated, with the help of a BigGAN algorithm (available via ArtBreeder), some historical and contemporary images of eyes and brains. The images had been drawn from the Wellcome Collection, a UK repository of medical images. The idea behind *Neuromatic* was to explore perception as a process unfolding between the eye, the brain, and the world—in humans and machines. The video's title is constructed from two terms: *neuro*, referring to the nervous system, and *matic*, standing for something pertaining to the eye (*mati* in modern Greek), but also sharing the meaning of the -matic suffix (as in *automatic*, "willing to perform"). *Neuromatic* aims to capture this link between the eye and the brain in the visual apparatus. It also invites a reflection on whether vision itself can be understood in machinic terms.

How to see better than humans

This chapter continues the interrogation of imaging and perception by focusing on the problem of machine vision, but it also analyzes the problem *with* machine vision. As part of this analysis, I will embark on an experiment in "conceptual engineering," an approach that links the pragmatism of the machine-building discipline with its Latin etymology in *ingeniare*, meaning to devise, create, or contrive—and thus also to play. With some help from select computer scientists and neuroscientists, I will explore the possibility of building a perception machine—of a nontrivial variety, no less. The chapter draws on two science papers, separated by a couple of decades, by authors who made significant contributions to the debate on the relationship between humans and machines: Heinz von Foerster's 1971 "Perception of the Future and the Future of Perception," which introduced the concept of a "non-trivial machine," and Gerald M. Edelman and George N. Reeke Jr.'s 1990 "Is It Possible to Construct a Perception Machine?" I will engage with those papers in an attempt to construct a conceptual scaffolding for a theory and praxis of machine perception, while interrogating the role of photography and other forms of imaging in industry efforts to get machines to "see."

Mindful of Andreas Broeckmann's indication that "machine" "is a complicated concept that, if used carelessly, troubles any thought about technology more than it clarifies,"[1] I want to explain and contextualize the key terms and concepts that shape my argument. The term "machine vision" refers to the systems engineering discipline that works on the automatic extraction of information from digital images to enable machines to perform tasks requiring human sight—and to do it faster and more efficiently than human sight could. Such tasks may include quality control, identification, verification, positioning, and measurement, and can be found in applications such as detection systems in self-driving cars, manufacturing, security, or space exploration. Machine vision systems rely on cameras with sensors, processing hardware, and software algorithms. Their software side is underpinned by a cognate discipline called "computer vision," a subfield of the broad area of artificial intelligence, which theorizes how the extraction of information from digital images actually occurs—although the two terms, "machine vision" and "computer vision," are sometimes used interchangeably. The primary goal of research in computer vision is to teach computers to see, and to make this seeing meaningful to them, at least within the parameters of the set tasks. In its current iteration, machine vision is aimed at imitating the way humans see the world, but this imitation attempt occurs only *after human visual processes have been redefined in computational terms*.[2] The ultimate goal of machine vision is to learn how to see *better*, that is faster and more efficiently, than humans.

The methodological premises for teaching computers to see were laid in the late 1950s; a 1959 paper by two neurophysiologists, David Hubel and Torsten Wiesel, titled "Receptive Fields of Single Neurons in the Cat's Striate Cortex," is widely referenced as foundational to the subsequent developments in computer vision. The paper was based on a series of experiments conducted by the two researchers on "lightly anaesthetized"[3] felines whose eyes had been "immobilized by continuous intravenous injection of succinylcholine"[4] to eliminate unpredictable eye movement. The cats were being shown pictures of various dots and light shapes, or were having such shapes projected onto their retinas, with a view to assessing the felines' brain activity and thus identifying cortical correlates of vision. While failing to detect any significant changes to the cats' neuronal activity as a result of being exposed to the light projections of various shapes, the experimenters had a breakthrough when some of the cats' cortical neurons started firing furiously in reaction to the cats being exposed to slits of light. Hubel and Wiesel eventually realized that what the cats' retinas were reacting to was not any specific shape, such as a dot or a line, but rather change in light intensity at the edge of a slide frame. The two researchers traced this activity to what they ended up calling "simple cells" in (what is now known as) the primary visual cortex. This experiment, conducted as it was on immobile nonhuman subjects, under laboratory conditions, led to the inauguration of one of the foundational assumptions—or, indeed, myths—of computer vision: the belief that the process of vision is multilayered and hierarchical, that it is possible to extricate the essence of vision in various animals (including those of a human variety), that the mechanism of edge perception is what lies at the core of vision, that the process is physiological and content-independent, and that machines can be taught to see "like humans" by mimicking this process of pattern perception at the level of pixels.

These formalist parameters were consolidated by a 1982 textbook of computer vision, simply titled *Vision*, by MIT neuroscientist David Marr. For Marr, vision was primarily a phenomenon of information processing.[5] "In this framework, the process of vision proceeds by constructing a set of representations, starting from a description of the input image, and culminating with a description of three-dimensional objects in the surrounding environment."[6] The framework assumes that the mechanism of primary visual processes such as edge detection or binocular vision is computational and that it works its way from those primary processes upward in its study of the neural circuitry supposedly enabling vision, all the way to the brain.[7] It is this understanding of "computer vision," and of its engineering counterpart, "machine vision," that I aim to critically interrogate here. My overall goal is to probe how, or indeed whether, machines can actually see at all. I am also interested in what it means for us humans to endow machines with the capacity for seeing, or rather, to classify *as seeing* their ability to differentiate between objects in the world on the basis of the light reflected off them and transmitted as data to those machines' processors. This critical investigation will allow me to explore the possibility of building a perception machine, on terms that engage with, but also go beyond, those delineated by biology and computer science.

From machine vision to machine perception

The shift from vision to perception enacted as part of my experiment is not entirely mine: in recent years many AI researchers have gone beyond the explicit ocularcentrism of machine vision to extend the study of machines' data capture operations to other senses, such as hearing, touch, and olfaction. Expanding their data sources from still images to sounds, music, and video, Google is now using the term "machine perception" in lieu of "machine vision." Indeed, this is the second most popular category, after "machine intelligence," in its database of AI research.[8] Object recognition premised on algorithms trained on processing large, partially labeled datasets using parallel computing clusters, originally trained on optical images but now expanded to sound, music, and video, forms the basis of its functioning.[9] The notion of machine perception shifts the focus from the flatness of 2D images to the 3D environment which produces them, and which hosts the seeing machine, yet it still retains the sense of predefined objects, with their corresponding categories.

The notion of "machine perception" I am proposing here, adopted in a wider sense than the one offered by Google, aims to address one of the key blind spots of computer vision today: its inability to fully account for how our brains actually work and how the translation process from retinal stimulation (which used to be known as a "retinal image," as discussed in chapter 3) through to the neural circuits of the brain occurs, while producing a sensation and an awareness of this sensation (recognized in the form of an image that we "see") in humans. It also aims to raise questions for the postulation of the discrete physiological unity called "the brain" as the core organ of perception. This is not just humanities-scholar talk, driven by imprecision and the bending of categories for the sake of intellectual delight. A well-known 1959 paper by cognitive scientist J. Y. Lettvin et al., "What the Frog's Eye Tells the Frog's Brain," demonstrated that perceptive activity that had been assumed to take place in the brain as a consequence of the retina being stimulated by light in fact had already begun in the eye. "The eye speaks to the brain in a language already highly organized and interpreted, instead of transmitting some more or less accurate copy of the distribution of light on the receptors," Lettvin and colleagues argued.[10] The exact location of perceptive processes and the exact working of their operative mechanisms remain difficult to pin down not just in frogs but also in humans. Machines, in turn, are designed to "interpret images very simply: as a series of pixels, each with their own set of color values."[11] My notion of machine perception aims to complicate the simplicity of this model—and to explore a different way of understanding what it might mean for machines to see.

Perception of the future

The "conceptual engineering" aspect of this chapter departs somewhat from the recent uses of this term in analytic philosophy, where it serves as a more programmatic and less playful thought device. David Chalmers, for example, defines conceptual engineering as "the process of designing, implementing, and evaluating concepts."[12] Fixing and hence stabilizing concepts is at least as important to him as the act of inventing them. For me, however, conceptual engineering has something of feminist cyborg bricolage about it: it is less permanent, less intent on fortifying edifices—and much more mischievous and punk. Bringing together the two aspects of Chalmers's proposition, "de novo engineering" and "re-engineering," it retraces the patterns and scaffoldings of the established ideas and texts to arrange some new, albeit temporary, constructions—while looking for safety exits.

As indicated earlier, there are good reasons for turning to the papers by von Foerster and by Edelman and Reeke, as these authors played an important role in redefining physiological phenomena such as perception and vision as programmable. As part of their research, they set out to redraw traditional disciplinary boundaries in an attempt to find a new way of understanding humans and machines, or even humans *as* machines. Von Foerster was an Austrian physicist and engineer with an active interest in biology who spent most of his working life in the United States. He was one of the key participants in the Macy conferences, a series of meetings that took place between 1940 and 1961 which led to the emergence of cybernetics as a transdisciplinary field studying biological and mechanical systems. Von Foerster gained recognition for his work on second-level cybernetics, i.e., the proposition that the system's feedback loop needed to incorporate the observer into its operations, with human self-reflexivity becoming part of the cross-systemic operations across different levels. Von Foerster was a prolific writer, often veering beyond the confines of his discipline or even beyond science as such to offer a wider commentary on society and the world. “Perception of the Future and the Future of Perception,” first presented at the opening of the Annual Conference on World Affairs at the University of Colorado, Boulder, in 1971, is an example of this approach.

The paper opens with a turn to systems theory as a way of thinking about sociocultural change, while also pointing out the limitations of the static model of the system, whereby change is seen as an aberration to be corrected rather than an incentive to perform alteration at the systemic level. Taking account of systemic operations, von Foerster goes on to seek an opening within the system, toward an execution of human freedom and agency—and it is perception that he turns to for this purpose. (We can see similar intimations in Flusser’s work, which was largely inspired by cybernetic thinking.) Specifically, von Foerster advises us that “if we wish to be subjects, rather than objects, what we see now, that is, our perception, must be foresight rather than hindsight.”[13] In other words, he calls on us not to fall back on the established patterns of seeing things but rather to follow unexpected routes and pathways, taking a lesson from children. So far, so pastoral. He is of course not the first thinker to evoke the idea of original untarnished vision as a supposedly better entry point into the truth of the world: Enlightenment philosophers advocating a return to the “state of nature,” romantic poets, and surrealist artists were there before him. Yet von Foerster has some interesting things to say about the current atrophy (or, as he terms it, “castration”) of perception, for which he blames the commodification of information and “an educational system that confuses the process of creating new processes with the dispensing of goods called ‘knowledge.’”[14] To offer a remedy for this state of affairs, von Foerster introduces the concept of a “non-trivial machine,” where the “machine” “refers to well-defined functional properties of an abstract entity rather than to an assembly of cogwheels, buttons and levers, although such assemblies may represent embodiments of these abstract functional entities.”[15]

While a “trivial machine” is premised on a one-to-one relationship between its “input” and its “output,” thus delivering predictable, indeed consistently identical, results (at least in theory), in a nontrivial machine “its previous steps determine its present reactions.”[16] While both systems are deterministic, the second one is unpredictable because its output changes with what it has picked up in the previous cycles of its operation. Von

Foerster does appreciate the reliable predictability of various trivial systems, such as toasters or cars, but raises concerns about the application of trivialization to other system—and this is where his concept of the machine and his argument become particularly interesting. He takes the example of a student, whom von Foerster sees as a potentially nontrivial machine, becoming completely "trivialized" when he (*sic*) enters a higher-level machine, i.e., a university, with its predictable teaching and uncreative testing. "Tests are devices to establish a measure of trivialization," von Foerster proclaims. "A perfect score in a test is indicative of perfect trivialization: the student is completely predictable and thus can be admitted into society. He will cause neither any surprises nor any trouble."[17] While the human (the student) is seen here as a machine, still in the process of being constructed in his relatively young age, the higher-level operations of the recursive loops enable the possibility of creativity within the system, made up of other machines (educators, forms of knowledge and ways of their transmission, pedagogies, buildings, and infrastructures)—but only if the human agents execute their potential: in other terms, when they embrace the incomputable as part of systemic rationality.[18]

It may seem obvious that educators or indeed most humans would want to do this, and that they would reach for the systemically determined yet ultimately undefined degree of freedom available to them. Yet, as confirmed both by von Foerster (who sees reluctance to exercise freedom as a form of illness) and Flusser (for whom we always operate under the condition of systemic predictability, be it on the level of the technical apparatus, society, or the inevitably entropic universe),[19] such acts of reaching out toward freedom, i.e., attempting to change the system's course of action, are quite rare. Von Foerster's paper ends with a rather humanist call to cherish the "troublemakers," whom he says we will recognize "by an act of creation: 'Let there be vision: and there was light.'"[20] Interestingly, though, it is through perception, i.e., the way of seeing attentively,[21] that we (as educators and assessors of students) are said to be able to introduce novelty into the input-output process and thus create a systemic opening. This could be a first step in attempting to build a nontrivial perception machine.

But how *do* we make students, or indeed anyone, see better, without knowing in advance what we want them to see, what kind of frameworks we are looking through and aiming for? Is there not a danger that this call on the part of von Foerster will ultimately sound banal, or in fact trivial:

at best vaguely humanist, at worst a recipe for all sorts of educational libertarians to break with the expertise and care developed within the educational system? Frequently described as a "polymath," von Foerster is one of a long line of (predominantly male) scientists who opine on wider societal issues, making quick excursions to scientific models and protocols in order to develop explanations of the world while demonstrating strange blind spots when it comes to the cultural aspects of the systems they write about, including the terminological and ideological determination of their concepts. (Of course, all conceptual frameworks are ideologically determined, but it is the unawareness or obfuscation of this phenomenon, coupled with the ascription of "ideology" to one's intellectual opponents, that I am highlighting as a problem here.)

This model of what we might term "truncated rationality" is still very much at work in the very hotbed of cybernetic and postcybernetic thinking: Silicon Valley. In his poignantly titled *What Tech Calls Thinking: An Inquiry into the Intellectual Bedrock of Silicon Valley*, Stanford literature professor Adrian Daub points to an "amnesia around the concepts that tech companies draw on to make public policy,"[22] coupled with the fetishization of the novelty of the problem. This approach is premised on the overlooking or even negation of the critical and analytic tools that had previously been used by others to deal with similar problems. Through their pronouncements and publicity campaigns, Silicon Valley entrepreneurs can position themselves as visionaries, while disenfranchising "all of the people with a long tradition of analyzing these problems—whether they're experts, activists, academics, union organizers, journalists, or politicians."[23] What tech calls thinking is thus a form of instrumental rationality which has been shaped by the Randian myth of individualism and the logic of capital.[24] It is not only profoundly anti-intellectual but also, contrary to its own self-image, counterrevolutionary. "This is where the limits of our thinking very quickly become the limits of our politics,"[25] adds Daub. But while Silicon Valley is its most visible actualization, this form of truncated rationality exists and indeed thrives in scholarly disciplines such as analytic philosophy, linguistics, biology, cognitive psychology, and neuroscience—a cluster of disciplines that provide intellectual foundations for the field of artificial intelligence, and which frequently rely on copious amounts of technological and military funding.

By way of responding to the omnipresent working of truncated rationality in our increasingly automated and datafied society, Luciana Parisi has come up with what she has termed the "Alien Hypothesis," a proposition which embraces "an abductive, constructionist, experimental envisioning of the working of logic . . . as part of a speculative image of the alien subject of AI."[26] What is interesting about her idea is that it comes *from within* the logical parameters of cybernetic thought. Her hypothesis is designed as an alternative to both the Cybernetic Hypothesis, positing that the only possible escape from the constraints of the computational system is a flight into zones of opacity and invisibility involving the total negation of the logic of the system, and the Accelerationist Hypothesis, which advocates dismantling the system's power from within by exhausting its operations. Instead, Parisi proposes to engage with the systemic operations on their own logical terms with a view to discovering "an alien space of reasoning"[27] *within* the system. Drawing on the consequences of mathematician Gregory Chaitin's identification of incomputables within computational systems,[28] she argues for the possibility of finding such a space *beneath* the scripted and looped servomechanics of the system that make it seem like there is no other way. "The challenge," she suggests "is to determine the kind of appropriation that can open political possibilities with technology from within its mode of existence."[29] Parisi's solution lies in reconceptualizing the medium of thought beyond its automation and instrumentalization within the current machinic systems. Mine here, in turn, involves taking a step back alongside our cognitive-sensory spectrum to offer perception rather than thought as a primary mode of engaging with the world, one through which such an attempt at undoing the system's operations is most likely to succeed.[30]

Vision beyond the brain

So, is it possible to construct a perception machine? This question was originally posed by biologist and Nobel laureate Gerald M. Edelman and his collaborator George N. Reeke Jr. in their 1990 paper. Edelman and Reeke's work was part of their wider project on the neglect of findings from evolutionary biology in AI research. In an article written two years prior, they chastised AI researchers for remaining too wedded to "epistemological assumptions drawn on the one hand from the arguments of Alan Turing and Alonzo Church about the universal problem-solving capabilities of computers (suggesting that the brain may be understood as a computer) and on the other hand from the reductionism of molecular biology (suggesting that the brain may be understood as a collection of units that exchange chemical signals)."[31] They also had some rather critical things to say about "neural network computing"[32] as an approach that promised to develop machine vision by modeling it on the operations of the human brain. For them "neural network computing" is a misnomer because the approach that underpins it is premised on a badly conceived analogy between neural networks in the brain and computer networks, and not on the way biology actually understands and works with neural structures. The strictly computational approach, they argued, cannot really tell us much about perception because of its foundational error—namely, the belief that "objects and events, categories and logic are given and that the nature of the task for the brain is to process information about the world with algorithms to arrive at conclusions leading to behavior. . . . This 'category problem' leads directly to the inability of AI systems to cope with the complexity and unpredictability of the real world."[33]

Recursive neural networks used in machine (aka deep) learning that consist of layers of nodes can be said to have partly addressed the problem of being unable to deal with the increasing complexity and uncertainty of the world. Indeed, they have had significant successes in identifying patterns and trends in imprecise data, as evident in applications such as face recognition, medical data analysis, weather prediction, or natural language translation. Yet neural networks as currently conceived in AI research still do not ultimately challenge the assumption "that information exists in the world," while the organism "is a *receiver* rather than a *creator* of criteria leading to information."[34] I am particularly interested in Edelman and Reeke's critique of the idea of objects and events existing in the world out there, to be seen, grasped, and manipulated by us. This critique corresponds to the philosophical position of "conscious realism" espoused by Donald D. Hoffman discussed in the previous chapter. This is not to say that there is *nothing* in the world, only that the preconceived and discretized objects that allegedly present themselves for our vision to capture are outcomes of a creative, dynamic yet ultimately undetermined process.[35] We see what we need to see, argues Hoffman. For him, the nature of this need is biological, or, more precisely, evolutionary, in the sense that the purpose of our seeing is our survival—and this is where his intellectual trajectory coincides with Edelman and Reeke's. Indeed, as a riposte to computational schematism in physics, Edelman and Reeke suggest that any viable theory of categorization and intelligence to be used in AI research needs to embrace the Darwinian model of selection, but adjusted for the working of the neurons of a single organism operating during its lifetime. This conclusion serves as the grounding for their 1990 paper about the possibility of constructing a perception machine.[36]

And, indeed, in the paper they propose such a construct. Named, without any equivocation or irony, Darwin III,[37] this machine's architecture is premised on their proposition that categorization (rather than the more straightforward object recognition as applied in computer vision systems) is a critical component of perceptual systems. Perception does involve "the adaptive discrimination of objects or events through one or more sensory modalities, separating them from the background and from other objects or events,"[38] but the difference here is that these objects are not predetermined. Instead, for Edelman and Reeke perception is not about grasping representations, as the categories used by the organism or machine do not exist in the world to be recognized by this organism or machine: rather, those categories are actively constructed (or, as Edelman and Reeke put it, they are meaningful to the perceiver "in a given situation").[39] Perception thus becomes redefined as "an active sensorimotor process, requiring exploration and depending on past experience."[40] In this model categorization needs to be seen as biological, not mathematical, because it cannot be solely expressed in symbolic form. The biological aspect of these organisms implies that movement through the world is needed for recognition, and hence perception and intelligence, to be enacted. This model is premised on the interesting idea, formulated in the authors' earlier paper, of the brain as "a selective system operating in somatic time,"[41] one which resonates with the phenomenological approach to AI design, which is increasing in popularity.

Yet Edelman and Reeke's concept of the perception machine itself remains hamstrung in this acultural model, with machinic operations positioned as primarily driven by natural selection. Cultural transmission of information is dismissed by them as not being relevant to the *evolutionary* development of working perceptual systems. Even though Darwin III is supposedly more firmly placed in the world, its orientation, vision, and goals are crippled by the temporality of its movement, with "somatic time" requiring an erasure of the said soma's singularity at the expense of the (supposedly) timeless operations of natural selection. There is therefore a danger that Edelman and Reeke's perception machine may end up being rather trivial, because the information it will pick up to develop its categories, even if collected in a more dynamic and embedded way, will by design have been stripped of any cultural specificities—which here only amount to noise. However, designating which aspects of the surroundings can be classified as nature and which belong to the realm of culture is not straightforward.[42]

Anthropologist Tim Ingold argues that the difference between nature and culture is just a matter of temporality: history (which manifests as "culture" in different epochs) simply operates on shorter time scales than evolution (which proceeds according to laws of "nature"). Ingold goes on to show how Darwin's *Origin of Species* challenged the eighteenth-century view of "man" as a being equipped with special characteristics such as reason and morality, characteristics which were meant to separate him from the other species. Darwin postulated that the difference between the human and other animals was in fact one of degree, not of kind, and that evolution, which he had originally termed "descent with modification," was the manner through which this change toward rational man, endowed with intellectual and moral faculties, actually occurred. Yet the problem with that view was that it inaugurated another differentiation which today we would (rightly) describe as racist: one between the savage and the civilized man. The only way out of this dilemma was to attribute "the movement of history to a process of culture that differs in kind, not degree, from the process of biological evolution," explains Ingold.[43] It was also to put forward the idea of two parallel kinds of inheritance in human populations: one biological (involving the transmission of genetic information encoded in the DNA and referring to the core identity of the human—e.g., walking), the other cultural (taking place through social learning—e.g., being able to play an instrument). Yet Ingold argues, following many other cultural anthropologists, that there is nothing "purely" natural about walking—and that neither is learning to play an instrument a fully cultural experience, separated from the transmission of embodiment. Instead, "those specific ways of acting, perceiving and knowing that we have been accustomed to call cultural are enfolded, in the course of ontogenetic development, into the constitution of the human organism," which makes them equally "facts of biology."[44] At the heart of the problem lies not so much the conflation of the biological with the genetic, Ingold points out. The decoupling of the processes of historical and biological change should therefore rather be treated as a detemporalization, a process premised on the fictitious positing of separate temporalities for certain kinds of changes over others. This fallacious model is still very much with us. It subtends the present consensus with regard to science as a method—and the disciplinary division between the sciences and the humanities.[45]

Machine vision and epistemic (in)justice

Ingold's argument, which builds on the work of psychologist Daniel Lehrman, poses a challenge to any attempt to construct perceptual systems that will be capable of passing on their evolutionary traits to their offspring but will not be able to pass on any cultural influences (or, indeed, that will need to be free from such influences). Any traits that *will* get passed on will always carry both "natural" and "cultural" inscriptions, in a manner that will not allow for their easy decoupling, but the construction of culture as a separate domain of uninheritable features will perpetuate the distinction, while allowing computer scientists to forgo embodied and embedded modes of perception and cognition. This disembodied model of computer vision results in the preservation of one of the biggest science (and computer science) myths: the belief that data bias understood as cultural bias, once eliminated, will result in data that is both pure and fair.[46] We are regularly presented with consequences of such essentialization of biology and "the brain," at the expense of "cultural traits," in cognitive and computer science. Two examples from 2020 include the video conferencing platform Zoom's background algorithm, which removed the head of a Black academic any time he tried to use a virtual background, and the Twitter cropping algorithm, which privileged the showing of white faces in cropped images in its timeline.

While the computer vision system reveals itself not to be particularly perceptive, the consequences of its racialized blind spots are anything but trivial. Indeed, the algorithms that run within it are the same ones that make decisions about people's social, financial, or legal status, including punitive action at border control, denial of credit, prediction of educational failure, or assignment of criminality. While the early Google image recognition algorithm was shown in 2015 to auto-tag pictures of Black people as "gorillas,"[47] there is an ongoing problem with face recognition of Black females, with the high false match rate justified by industry experts as being the result of a combination of the difficulty of lighting a Black face and the makeup worn. Denying explicit bias, Thorsten Thies, director of algorithm development of the German company Cognitec which supplies facial recognition systems to governments, explained in a troublingly disarming manner that it is "harder to take a good picture of a person with dark skin than it is for a white person."[48] One factor is that the image databases that serve as training sets for the algorithms are not properly representative, being skewed, in terms of volume and quality, toward photographs of white males. But there is a deeper logic at work here, endemic to the whole systemic infrastructure involved in the production of cameras, lighting systems, image-processing software, and the visual and cultural training of photographers and image technicians, that produces a particular set of internalized norms. These norms that can then be presented as posing an "objective" difficulty in taking a photo of a person with a dark skin. This mode of thinking, embedded in all sorts of technologies that precede the digital, is what Safia Noble has critiqued in her book *Algorithms of Oppression*. Noble's book analyzes the hidden operations of capital, race, and gender in algorithmic infrastructures, a process she terms "technological redlining."[49] Suggesting that "artificial intelligence will become a major human rights issue in the twenty-first century,"[50] she calls on us all to understand the architecture and logic of algorithmic decision-making tools in masking and deepening social inequality. As part of her analysis, Noble puts to rest the belief that unfair systemic decisions are just occasional aberrations which can be easily eliminated for the functionality and efficiency of the supposedly neutral system to be restored.

Murad Khan and Shinji Toya's project *Identity Turns on a Pixel When the World Becomes a Picture* poignantly illustrates the logic behind racial categorizations in machine learning and computer vision systems, an (il)logic that is premised on arbitrary categorizations replicating entrenched sociocultural perceptions of identity as only skin- (or hair-)deep. For their project the artists had scanned a photo of a young male, which was labeled by a facial recognition algorithm of a commercial software package as "Asian." They then morphed the photo in an AI-based image manipulation program, via the addition of facial hair and a cropped haircut, to create an image of a person that got labeled as White. Subsequently they analyzed the in-between stages in the morphed image to identify the threshold image over which the "crossing" between the racial categories occurred. It turned out the difference between the two images boiled down to a single pixel. "The minute, discrete, and tangible nature of the pixel as a racialized limit point between categories points to a social problem at the heart of abstraction in machine learning systems,"[51] they concluded, revealing a gap between human vision, which is premised on an indeterminacy to be negotiated, over and over again, as part of our lived experience, and its machinic counterpart—where such indeterminacy must be eliminated at all cost.

CDO at Twitter Dantley @dantley responded to the 2020 cropping algorithm's debacle discussed earlier with the chest thumping yet predictable: "It's 100% our fault. No one should say otherwise. Now the next step is fixing it."[52] Yet Noble is very clear that "algorithmic oppression is not just a glitch in the system but, rather, is fundamental to the operating system of the web."[53] Mitra Azar, Geoff Cox, and Leonardo Impett similarly suggest that "in a structurally unequal society, it is exceedingly difficult to make a 'fair' algorithm; and it is effectively impossible to make an algorithm which is both fair and effective. . . . In a society which is unfair, a classification-machine will always be unfair (in at least one sense)."[54] It is therefore not enough to de-bias the data. Instead, we need to ask bigger questions about the forms of injustice embedded in the systems that host it. We also have to ask what it means when the elimination of the glitch, while desirable from a technical point of view, ends up making the punitive surveillance running on this data even more efficient. The correction of the data bias does not correct the violently penetrative and extractivist logic of the computer vision system: it actually strengthens it. We could thus say that we need to identify the unjust operations of the nontrivial *vision* machine to start thinking of building a nontrivial *perception* machine.

With this, as part of the "conceptual engineering" project outlined at the outset of this chapter, we are now shifting toward a conceptual expansion of the notion of the machine. In line with its cybernetic legacy from von Foerster and colleagues, the term "machine" departs from its strict engineering connotations to embrace any kind of system, be it mechanical or biological, of varying degrees of complexity. In their *Autopoiesis and Cognition*, a key text of systems theory which paved the way for applying systemic thinking not only to biological organisms but also to forms of social organization,[55] Humberto Maturana and Francisco Varela defined the machine as "a unity in the physical space defined by its organization, which connotes a non-animistic outlook, and whose dynamisms [*sic*] is apparent."[56] In this framework, systems are arranged in a nested manner, from the microscopic to the cosmic, while undergoing internal transformations or even cross-systemic mutations. In the 1960s and 1970s the cybernetically inflected notion of the machine became a potent concept for philosophers and cultural theorists attempting to articulate different levels of sociopolitical complexity while taking into account the biological and technical constitution of both individuals and societies. We can mention here Gilles Deleuze and Félix Guattari's war and desiring machines, Michel Foucault's dispositive, and Vilém Flusser's apparatus. Building on this legacy, "the perception machine" embraces multiple, albeit interwoven, levels of meaning, from the biological level of the perception apparatus in human and nonhuman animals through the technical image-making apparatus, all the way to the sociopolitical setup—and to its technical infrastructures which have a planetary reach.

By now it should have become clear that in my attempt to build a nontrivial perception machine I am not really constructing a device, at least not in any straightforward way, but rather a conceptual framework, a scaffolding for both framing what we term the world and seeing this world better. Such a perception machine will have to do more than just identify symbols, avoid bias, or even account for the environment. In its architecture the idea of the brain as a discrete perceptive organ will need to give way to a dynamic interaction between the organism, with its constantly changing embodiment, and the environment—which, in turn, is not a constant but which "exists only in relation to the organisms that inhabit it, and embodies a history of interactions with them."[57] Indeed, there is no perception machine outside of its cultural and historical embeddedness. This acknowledgment calls for "nothing less than a new approach to evolution," one that does not posit "a radical break between evolution and history, or between biological and cultural evolution" but rather sees history as a "continuation into the field of human relations of a process that is going on throughout the organic world."[58] The very gesture of embarking on the task of trying to build a nontrivial perception machine is also meant to serve as a multiscalar attempt to both rethink our human perception and vision, and challenge the parameters of the emergent vision machine in our globally networked world. (This approach sums up the rationale behind my "conceptual engineering" project.)

The parameters of such a vision machine have been poignantly analyzed by Virilio in his 1994 book. Initiated with the development of the prostheses of sight such as the telescope in 1608 but accelerated three centuries later with the proliferation of “seeing machines,” the transformation of vision is understood by Virilio in terms of a linear shift from the organic to the mechanical. It also entails an incontrovertible loss. Virilio’s vision machine thus ends up being posited as “an autonomous technological system”[59] which disturbs the original organic unity and purity of human sight. John Johnston observes that “in Virilio’s theory there is no positive side to the deregulation of perception (unlike Deleuze and Guattari’s deterritorialization), no positive value, aesthetic or otherwise, to the freeing of perception from preestablished codes.”[60]

Drawing on some of Virilio’s concepts, as well as his political intimation with regard to the unjust and immoral uses of the visual technologies of surveillance, regulation, and control, in this book I go beyond his humanist melancholia to embrace an understanding of machines that encapsulates their organic and nonorganic iterations, that changes over time, and that constitutively shapes the human sensorium in different ways, while also enabling its reframing and recoding. I am doing this with a conviction that it is only through an acknowledgment of our own technical heritage, and through a prudent and responsible engagement with machines as co-constitutive of our bodies, societies, and environments, that we will be able to design better ways of being in the world—to develop an ethics and politics for a humachine world that is not just a form of organicist moralism. But, to do this, we will need not just a new ethical framework but also a new optics.

A nontrivial perception machine must be antiracist

The transformation of vision acquired a new intensity in the age of what Shoshana Zuboff has termed “surveillance capitalism.”[61] For Zuboff the term describes a novel economic order, one that is characterized by parasitism and attempts at behavior modification, whereby human experience is reduced to raw material for the unprecedented operations of power wielded by technological companies. She defines this development “as a coup from above: an overthrow of the people’s sovereignty.”[62] While Zuboff offers some insightful analyses of the working philosophies and practices of companies such as Apple, Google, Amazon, Microsoft, and

Facebook, her proposed solutions are arguably less discerning. Leif Wetherby critiques Zuboff's study for being premised on a rather empty promise of systemic reform from inside. Wetherby argues that the metaphor of an all-consuming surveillance machine her book embraces is too totalizing, and that this machine "is really a dispersed and uneven set of global infrastructures."[63] "This system does not 'see' so much as it captures," he suggests, with the visual perception of our bodies being intertwined with the incessant seizure, processing, and distribution of abstract data about us, on multiple levels and in multiple domains.[64] The current vision machine we are living in is thus both a surveillance machine *and* a data capture device. It is also, as rightly pointed out by Virilio, a war machine, serving as it does as a digitized battleground featuring multiple operations of capture *and* carnage, with real-life consequences for human and nonhuman lives. Yet it is also, we must not forget, a pleasure machine, with photography and other forms of automated image-making shaping the flows of desire of human users who cannot be reduced to mere cultural "dupes."

My attempt to build a nontrivial perception machine presented here is thus aligned with Bernard Stiegler's analysis of automation in pharmacological terms,[65] where the process, although harmful to our individual and social life—as evidenced in our overall "ill-being," "nihilism," and "humanity's doubt about its future"[66]—can be worked through to release the machine's curative properties. The nontrivial perception machine we are building here will need to be able to scan through the obscure logic of the surveillance machine, which is also a data apparatus and a mechanism of warfare. To do this, such a nontrivial machine would need to do more than just be neutral or unbiased (although it should be that too): it must also be decisively antiracist. We could go even further and suggest that, to counter both the racist legacy of the war machine and the capitalist extractivism that fuels it, it must also, following Ariella Aïsha Azoulay, be counterimperialist. In her magisterial volume *Potential History: Unlearning Imperialism* Azoulay challenges the regime of rights and privileges that has shaped photography and the modes of perception and representation inaugurated by it since its early days, being designed as a medium through which "the world is made to be exhibited." The world is thus offered "only for a select audience."[67] She goes so far as to claim that "Photography developed with imperialism; the camera made visible and acceptable world destruction and legitimated the world's reconstruction on imperial terms."[68]

It may seem at first glance that Azoulay's argument is only aimed at humancentric photography, which is meant to be displayed and seen by (select groups of) humans—and not at large datasets feeding and shaping machine vision today. Yet the same form of rationality arguably underpins the production, storage, and categorization of all photographic and para-photographic images (including those rendered by CGI or GANs) because their constitutive logic, history, and modes of framing still hark back to the imperial mindset that legitimated classification as supposedly neutral, while putting it to work with Empire's goals in mind. The location and format of imperial rule have shifted today: in their eponymous book Michael Hardt and Antonio Negri argue that Empire now has no specific boundaries or territorial center of power. It has become "a *decentered* and *deterritorializing* apparatus of rule,"[69] with global capital flows enacting a form of biopolitics by being involved in the production of social life itself as an overlapping nexus of economic, political, and cultural forces. The neo-imperial war machine is thus first and foremost a hegemonic surveillance network: it conquers by implicit consent and by the scale and invisibility of its penetration. All this is not to say that no nodes of power's concentration can be identified within this new imperial apparatus. However, the increasing shift of domination and decision-making from governments to corporations, from Washington to Silicon Valley, from the United States to China, and from humans to algorithms creates an uncertain and fuzzy geopolitics in which political and technological black boxes obscure the location of power as well as its actual operations. It also shifts the onus of responsibility onto individual citizens—whose primary yet already dual identity in this system is that of both network users and data points. The neoimperial surveillance machine is thus an updated version of Virilio's vision machine. A nontrivial perception machine could be seen as its conceptual and technical counterpart, one that offers an opening into a new vision of both ourselves and the world.

The Recognition Machine

The Recognition Machine by artists Antje Van Wichelen and SICV (Michael Murtaugh and Nicolas Malevé) can be seen as one possible enactment of such a nontrivial perception machine that is also actively counterimperialist. I had an opportunity to interact with the version of the work presented

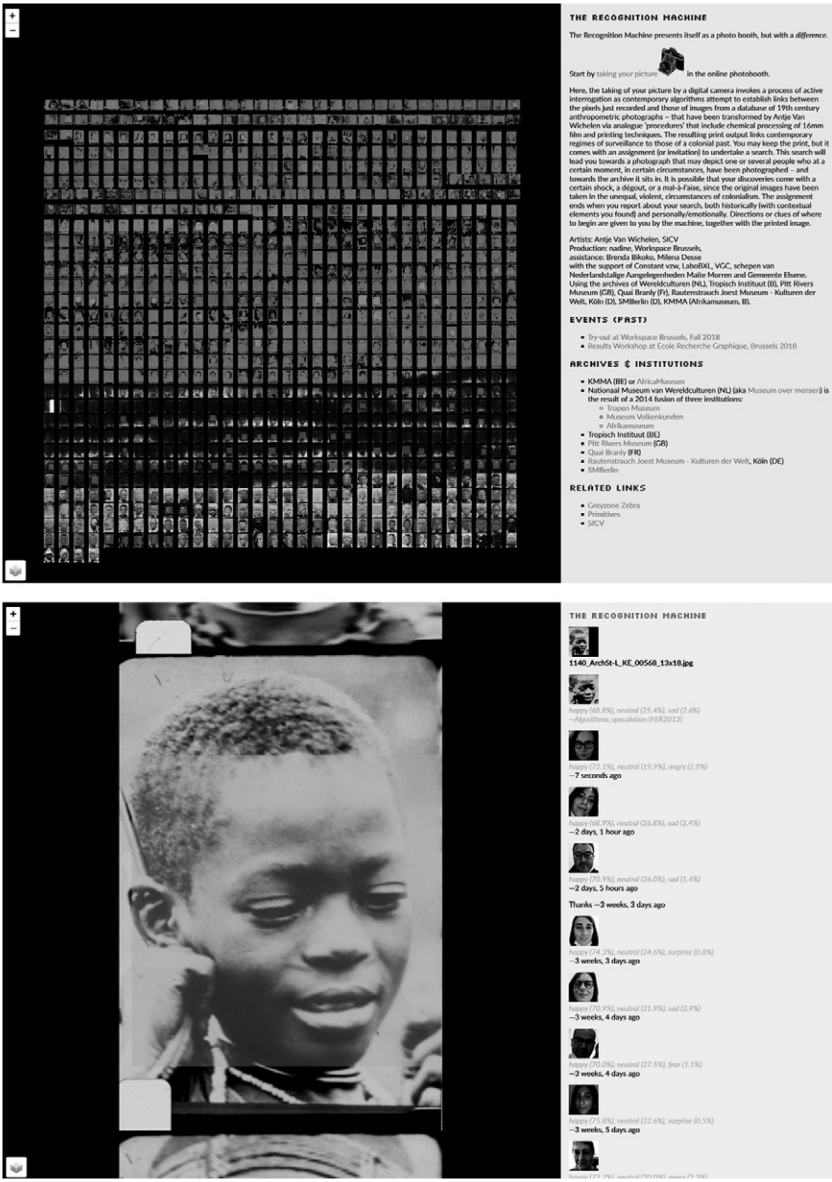

Figure 4.2
Antje Van Wichelen and SICV, *The Recognition Machine*, 2018.

at the Photoszene Cologne festival in May 2019, but the project also has an online counterpart. Looking like a photo booth, *The Recognition Machine* invites gallery visitors to enter and take a digital photo of themselves. The act of taking a photo activates an algorithm that attempts "to establish links between the pixels just recorded and those of images from a database of 19th century anthropometric photographs," which have been transformed by analog techniques. "The resulting print output links contemporary regimes of surveillance to those of a colonial past."[70] The link between the images pivots around the emotions identified by the algorithm in the viewer's face and linked with the emotions read in the archival photos. The reading was obtained by training the algorithm on the FER-2013 dataset, in which each image had been assigned one of seven emotions: anger, disgust, fear, happiness, sadness, surprise, or neutral. Any possible misrecognition of emotions that occurs as part of the process serves as an alert to the system of consequences that predictive technology is imbricated with: while labeling here is just an innocent game for art audiences, the misrecognition of image links, their wrong categorization and ascription, has serious consequences for the lived lives of many. The visitor may keep the print obtained, but they are also asked to explore further the posited analogy and thus go deeper both into the archive and the colonial history of portraiture. Demonstrating the pixel-level slide between races as enacted by the logic of computer vision systems analyzed by Murad Khan and Shinji Toya, *The Recognition Machine* also shows us that all images exist as part of the imperial-colonial network of visuality, a network that renders some bodies as visible and proper while deeming others as illegible or illegal. What is particularly interesting about this project is that the artists dispense with the idea of a singular image as a stand-alone artifact to be consumed, classified, and otherwise exploited, showing that *all* images are part of multiple networks of knowledge and data exchange. *The Recognition Machine* thus offers a model of the perception machine as an invitation to study the production of visuality, the image networks, and their infrastructures; their underlying data and databases; the algorithms that shape both their production and their networking.

We could therefore conclude that a *nontrivial* perception machine would need to encourage an ethicopolitical engagement with images, their histories, databases, and infrastructures. It should also entail strategies for entering the database on the part of the human, with a view to deindustrializing visuality and vision. It is therefore not just a matter of seeing what is inside the archival machine and how "it" thinks but also of creating conditions for thinking about human and machine vision otherwise. The human may not be able to see all the available images contained in multiple databases and data clouds, trace all the possible connections between them, or take cognizance of all the categories and labels on offer. But what is possible—and indeed imperative—is for the human viewer to take stock of the logic of opacity and scale that shapes the AI-driven perception machine, to ask questions about its operations and to demand a better (fairer, more historical, more explicitly antiracist and counterimperialist) engagement with the image and data flow.

Unlearning imperialism, as Azoulay points out, needs to involve attending to "the conceptual origins of imperial violence, the violence that presumes people and worlds as raw material, as always already imperial resources."[71] The fuzzy borders of today's Empire, coupled with the individual benefits of becoming-data for platform capitalism's servotechnology, make it easier for this form of violence to be seen as consensual—or not as violence at all. Taking first steps toward building a nontrivial perception machine can help us see this form of extractive biopolitical violence for what it is—and then start devising operations for countering it. A nontrivial perception machine can also be mobilized to render visible an envisioning process that operates as a condition of speculative thought.[72] We need to approach this task with the realization that, just as there is no space outside technicity (for us humans), there is no stepping outside visuality either. It goes without saying that such a machine cannot be built once and for all: it will need to be regularly rebuilt, retuned, and reprogrammed.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in ITC Stone Serif Std and ITC Stone Sans Std by New Best-set Typesetters Ltd.

Library of Congress Cataloging-in-Publication Data

Names: Zylinska, Joanna, 1971- author.
Title: The perception machine : our photographic future between the eye and AI / Joanna Zylinska.
Description: Cambridge, Massachusetts : The MIT Press, [2023] | Includes bibliographical references and index.
Identifiers: LCCN 2022050332 (print) | LCCN 2022050333 (ebook) | ISBN 9780262546836 (paperback) | ISBN 9780262376624 (epub) | ISBN 9780262376631 (pdf)
Subjects: LCSH: Photography—Philosophy. | Digital images—Social aspects. | Image processing—Digital techniques. | Artificial intelligence.
Classification: LCC TR183 .Z955 2023 (print) | LCC TR183 (ebook) | DDC 770.1—dc23/eng/20230405
LC record available at https://lccn.loc.gov/2022050332
LC ebook record available at https://lccn.loc.gov/2022050333

10 9 8 7 6 5 4 3 2 1