University of Bergen

Department of Linguistic, Literary and Aesthetic studies

KUN350

Master's Thesis in Art History

Spring 2021

Patterns of Discrimination

On Photographic Portraits as Documents of Truth in Automated Facial Recognition

Tuva Mossin

Sammendrag/Abstract

Denne avhandlingen tar for seg fotografiers rolle i treningen av ansiktsgjenkjenningsalgoritmer, samt i selve den tekniske prosessen hvor ansikter analyseres. Gjennom en lesning av tre ulike kunstprosjekter som på ulike måter anvender eksisterende ansiktsgjenkjenningsteknologi til å problematisere denne praksisen, etablerer jeg hvordan ulike fordommer – særlig hva angår fotografiets status som objektiv representasjon av verden – påvirker systemenes evne til å analysere ansikter. De aktuelle prosjektene er *ImageNet Roulette* (2019) av Trevor Paglen og AI-forsker Kate Crawford, *How do you see me?* (2019) av Heather Dewey-Hagborg, og *Spirit is a Bone* (2013-15) av kunstner-duoen Broomberg & Chanarin. Problemstillingen som oppgaven forsøker å besvare er som følger: hva kan disse kunstprosjektene fortelle publikum om ansiktsgjenkjenningsteknologi som praksis, og hvilken rolle spiller digitalt fotografi som slike systemers bindeledd til den analoge verden «utenfor» dem selv? Som svar på dette tar avhandlingen for seg selve den tekniske arkitekturen og hvordan den legger føringer for ansiktsgjenkjenningssystemers operasjoner alt i designprosessen. I tillegg diskuteres ansiktsgjenkjenning fra et historisk perspektiv, hvor forsøk på å knytte juridisk identitet til kroppen gjennom fotografi spores helt tilbake til mediets oppfinnelse på 1800-tallet.

This thesis is concerned with the role of photography in the training of automated facial recognition algorithms, as well as in the technical process of analyzing faces itself. By reading three different art projects which in various ways have appropriated existing facial recognition technology in order to scrutinize this practice, I establish how various biases – especially regarding the status of photography as an objective representation of the world – inform such systems' ability to analyze faces. The works in question are *ImageNet Roulette* (2019) by Trevor Paglen and AI researcher Kate Crawford, *How do you see me?* (2019) by Heather Dewey-Hagborg, and *Spirit is a Bone* (2013-15) by the artist-duo Broomberg & Chanarin. The central question which the thesis is trying to answer is as follows: what can these art projects teach the audience about facial recognition technology as a practice, and what role does digital photography play as such systems' connection to the analogous world "outside" of the systems themselves? Answering this, the thesis takes issue with the technical architecture informing facial recognition systems already at the level of design. It also discusses facial recognition from a historical perspective, tracing the endeavor to anchor legal identity to the body through photography all the way back to the medium's invention in the 1800s.

Patterns of Discrimination

On Photographic Portraits as Documents of Truth in Automated Facial Recognition

2021

Patterns of Discrimination: On Photographic Portraits as Documents of Truth in Automated Facial Recognition

Tuva Mossin

https://bora.uib.no/

Table of Contents

Acknowledgements

Above anyone else, I am hugely grateful for all the help and motivation from my dear supervisor, Prof. Sigrid Lien. The devotion you show towards your students is remarkable. Never could I have hoped for a more supporting and enthusiastic person to guide me through the overwhelming process of writing my first work at this scale. I do not know how I would have managed without you. Thank you.

Next, I wish to thank Post. Doc. Tonje Haugland Sørensen for your encouragement at an early stage of developing my thesis project. Not only was it you who introduced me to Heather Dewey-Hagborg's work, but your early reading tips have also been of great help in navigating the ocean of scholarship surrounding photography, surveillance and digital art.

Thanks also to Prof. Asbjørn Grønstad and the research group Media Aesthetics, who were kind enough to invite me to their seminar on the aesthetics of surveillance when I was only an MA-student in my first semester. This gave me much needed motivation as well as helpful insights into the interdisciplinary field of surveillance studies. I appreciate it.

I would also like to show my gratitude to those who have supported my work economically. Thanks therefore to Meltzerfondet, Bergen kommunes utdanningslegat and Benjamin Johannessens familielegat for generous stipends which have made the process of writing this thesis considerably less stressful.

Lastly, my family and friends deserve thanks for both emotional and practical support throughout the whole process. I am especially grateful for my parents' willingness to read through and comment this whole text. I also thank my dearest Olav for always giving me support and unconditional love, which not even a global pandemic has managed to affect. I could never have done this without you by my side.

“The good news about computers is that they do what you tell them to do.
The bad news is that they do what you tell them to do.”
- *Ted Nelson*

Introduction

It's a late-November day in Paris and I'm at the Centre Pompidou, having just wrapped up a seminar I've attended. With a couple of hours on my hands before the flight back to Bergen, I head straight for the section exhibiting early modernist art. Just like my fellow museum-visitors (most of whom are walking phone-first from work to work), I feel an urge to document what I see. My phone unpocketed, it's only a matter of minutes before I notice something curious: its facial detection function isn't limited to *real* faces, but detects *painted* faces as well, marking them with yellow circles on the screen. My interest piqued, I decide to conduct a little experiment: holding my phone up in front of me with the camera turned on, I move from work to work, curious to see if there's any kind of pattern regarding which kind of painted faces it can or cannot detect – and whether there are any notable differences between how the program reacts to various styles or levels of abstraction. What I find is something highly illustrative of the topic of this thesis. While the phone has no problem recognizing even some of the most abstract or blurry ("pixelated") faces painted in, for example, an impressionistic manner, there is one style that seems to give the camera function an extra hard time: the cubist portraits of Picasso.

On the face of it, this simple observation does not seem too surprising: the deconstructed figures in some of Picasso's paintings are so deformed, the perspectives so twisted and multifaceted, that no one would expect even the most accomplished system to be able to recognize the assemblages of seemingly random forms and colors as people – let alone to detect a face in the midst of it. Yet, *I could*. To the *human* beholder, there is no doubt as to what Picasso's motifs are meant to represent, even though the works' symbolic meanings are open for debate. Even if you dislike the artist's style and way of representation, the indexical meanings of his paintings are still obvious. This takes us to the core of the subject matter of this thesis, which is to analyze systems of automated facial recognition (AFR). More precisely, I will investigate AFR technology as a practice through a reading of three art projects which in different ways have appropriated tools from such technologies to produce portraits. I use the term *automated facial recognition* to denote any digital process of connecting faces to legal identities via photographic images or image flows. It goes from this definition that AFR comes in many forms in that such systems usually consist of a computer program working in tandem with a set of cameras through which people's bodies are registered. It is important to note, however, that not all optically based surveillance systems include facial recognition software.

The projects in question are, respectively, *ImageNet Roulette* (2019) by Trevor Paglen and AI researcher Kate Crawford, *How do you see me?* (2019) by Heather Dewey-Hagborg, and *Spirit is a Bone* (2013-15) by the recently separated artist duo Broomberg & Chanarin. Even though the forms and aesthetic expression of these works vary – as well as their particular relation to the technology in question – they all have in common that they, through different sets of artistic techniques, engender a critical "look back" at the technologies with which they were made. As all of these artists seem fond of mentioning, facial recognition is everywhere. While mainstream media usually reserve critiques of AFR to only the most obviously problematic uses – the most chilling examples can be found in China within the so-called social-credit system, as well the segregationist monitoring of Uyghurs in the Xinjiang province – there exist many more subtle examples that relate more closely to a Western public's daily experience. For my own part, the first conscious encounter with facial recognition was on Facebook many years ago, when the site suddenly started suggesting people for me to tag in the photos I had uploaded. Another common use for AFR is in smartphones, of which most new models offer the possibility of unlocking the phone and apps by recognizing the owner's face through the front camera. Apps such as TikTok and Snapchat, where face filters are very popular, are other examples. In the latter case, however, it is only talk of facial *detection*, which is part of, but not the same as, facial *recognition*.

The above are all examples of AFR in private or relatively limited social contexts, but even more is at stake when biometric technologies (of which AFR is only one example) are used in public spaces. A Norwegian reader is probably only used to encountering public AFR at airports, as these constitute border areas with a particular concern for security. One does not have to travel further than London, however, to find oneself in one of the world's most heavily surveilled cities in terms of number of CCTV cameras (Bischoff 2021). In light of how increasingly ingrained in our daily lives AFR technologies are becoming, the critical perspectives offered by the artworks in question are a welcome way of investigating this developing relationship between humans and machines. Importantly, by building on the rich tradition for the study of visuality and optics within art history, the works facilitate an embodied and historicized relation towards AFR not easily achieved by other means.

Scope and Structure

The question guiding this thesis is simple: what can the three art projects discussed here teach the audience about the AFR technologies with which they were made? As AFR's relation to the world "outside" the system is mediated through cameras and other optical devices, it is clear

that they rely on an idea of the photographic portrait as a document of truth concerning the depicted individuals. A secondary focus will thus be to question the role played by photography in AFR technology. In the process, a whole subset of related questions arises, ranging from the nature of identity and the idea of positive identification, to global and national power relations, to claims of mechanical objectivity. These will all be discussed separately along the way, whenever they appear particularly relevant. Regarding structure, I have organized the thesis as three case studies where each art project will be thoroughly analyzed in a separate chapter. A consequence of this strategy might be that some ideas are left behind when moving on to the next chapter. However, the central ones will be taken up again in the fourth and final chapter, which is constructed as a meta-analysis discussing the most intriguing finds in more detail.

The first chapter is focused on the work *ImageNet Roulette* by Trevor Paglen and AI researcher Kate Crawford, which is a multifaceted project existing in several versions. The first instance through which an audience could experience the work was as an online *app* published on a webpage in September 2019 to which people were invited to upload photos in order to have them categorized by a facial recognition algorithm. The app was later joined by an online essay contextualizing it to make the political dimensions of the work explicit. In my reading of this essay, it becomes clear that the algorithm had been trained in a database called ImageNet, which is the biggest and most used platform for training AFR systems. As the app turned out to be severely biased – especially along racial lines – the essay explains how this is a result of the highly dubious taxonomy informing the database in which it was trained. In this way, the project served as a platform through which to call attention to the role played by training databases in informing the "worldview" of algorithms, and also to raise questions about the possibility of building such a system that is politically neutral.

Shortly after the app's release, Paglen and Crawford opened a related exhibition in Milan titled *Training Humans*, in which a general evolution of facial recognition research was presented to the audience. Discussing this project – which I from here on refer to collectively as ImageNet Roulette (without the italics) – another set of questions arises. Here, it becomes clear how photographs play a crucial role in training AFR algorithms by serving as visual representations of the given categories. Questions regarding photography's status as a mediator between the algorithm and the world outside of the digital realm are therefore pertinent. The biases which will be discussed in the reception analysis of the app also spur questions of what constitutes "fair" use of AFR, and who has the right to determine this. Because Chapter 1 serves a double role as both a presentation of ImageNet Roulette and also as a general introduction to AFR, I will move on towards the end to situate the findings in a broader context of AI research.

Here, I problematize what within the field is being referred to as the "alchemy problem," which has to do with how AI research, according to some, finds itself at a crisis in terms of its scientific status. Finally, I ask whether the kind of artistic scrutiny exemplified by ImageNet Roulette can be read into a larger tendency within the field of visual studies, where scholars are starting to discuss networked digital photography in terms of what is referred to as a *viewing ecology*; a discussion which I will return to again in Chapter 4.

Chapter 2 is constructed as an analysis of the work *How do you see me?* by Heather-Dewey Hagborg. This project also saw the light of day in September 2019 when it was exhibited at the Photographer's Gallery in London. This is a digital video work presented as a series of images based on an experiment conducted by the artist. Seeking to determine the limits of what could be recognized as a face by a facial recognition algorithm, Dewey-Hagborg used what is called *adversarial processes* to "fool" the system, revealing how it does not at all understand what a face actually *is*. Compared to ImageNet Roulette, which can be said to center around the material and epistemic bases of AFR's *production*, *How do you see me?*, on its side, appears more concerned with the technology as a mode of *perception*. The result is a series of black and white oval shapes which to the human beholder look nothing like faces at all. This not only raises questions of visual representation in general, but also about what constitutes *recognition*.

Discussing this, I will take a step back to investigate the technical basis of AFR systems in order to make explicit how these constitute abstractive processes that turn information into increasingly compressed versions of the original input photos. While this step might first appear unnecessarily technical, the discussion is of importance in understanding both the relation between Dewey-Hagborg's face and the abstractions in *How do you see me?*, and also because it serves to give the reader a fundamental understanding of how AFR systems are essentially statistical probability calculation devices. While the political implications of this last point might not be crystal clear at this point, it is of central importance to the discussion in Chapter 4, where I will discuss this aspect of AFR in detail. Moreover, the work opens a broader discussion of the nature of intelligence implied by the term *artificial intelligence* (AI), in terms of which AFR systems are commonly described. Discussing this, I will draw on the work of philosopher Hubert L. Dreyfus, whose famous critique of AI from the 1970s still appears highly relevant. In light of this discussion, it becomes clear how idealist notions of form as separable from content have informed the field of AI since its inception, and how this constitutes serious problems when faced with semantic ambiguities – of which photographs are full.

In Chapter 3, I analyze *Spirit is a Bone* by the artist duo Broomberg & Chanarin. This work exists in several versions, including a photobook published in 2015 and various

installations exhibited across the world. In addition to the book, the basis for my discussion will be an installation encountered by me at the Hamburger Kunsthalle in 2018. Intrigued by this particular installation, I went on to order the book as well. It turned out that the series of eerie portraits were made with a facial recognition system specifically designed to photograph people without consent. This is in itself, of course, an interesting aspect of the work, but the artistic methods used for presenting the images also invite a lot of other reflections. While the installation positions its subject matter clearly within the digital realm, raising questions concerning the nature of body as a site for digital surveillance, the book, on its side, more clearly connects the use of contemporary AFR technology to a history of surveillance based on physiognomic measurement of the face. This is achieved by appropriating the categories from the twentieth century German photographer August Sander's oeuvre. Drawing mostly on the work of photography theorist Alan Sekula, I discuss how Sander's practice was based on dubious ideas surrounding the face as an icon of personal character. In this way, *Spirit is a Bone* opens up a discussion of the photographic portrait's role in pseudo-scientific practices with the human skull as their object of study, dating all the way back to the medium's invention.

The fourth and final chapter of the thesis is constructed as a meta-analysis of the preceding discussions in two parts. In the first, I summarize the general findings, placing them more specifically within a broader historical context. I am here especially concerned with how the integration of the photograph into disciplines such as craniometry and phrenology coincided with the rise of the modern surveillance state. Focusing on the deterministic nature shared between phrenology and present-day surveillance systems, I trace some common threads between the two, arguing that contemporary AFR technology can be understood as the latest "peak" in a centuries old evolution of practices concerned with anchoring legal identity to the body. Having established this, I go on in part two to analyze the artistic methods used for bringing forth the above reflections. Here, I am particularly concerned with how all three works play on contested ideas about the materiality of the digital realm in order to position themselves within a discourse on digital phenomena. Moreover, I compare the aesthetic expression of ImageNet Roulette, *How do you see me?* and *Spirit is a Bone* to the Berlin Dada group's use of so-called photomontage to ask if there is a special political potential in the photographic portrait when re-assembled and recontextualized.

What has not been noted until now is that all the works presented in this thesis are portraits of some kind and therefore belong to a genre with special connotations. As art historian Shearer West has noted "it is fair to say that portraits are — and have always been — used for documentary purposes" (2004: 57). From ancient times, portraits in various forms have been

used to document things as different as the mere existence of a person, to their power, to central events in a person's life. However, the imaginative and interpretative aspects of the genre, West argues, make portraits resistant to what she calls *documentary reductivism* (2004: 59). As the reader will be made explicitly aware, this is nevertheless exactly how photographic portraits are used in AFR. At the heart of the discussion throughout this thesis thus lies a tension between the portrait as *likeness* – documenting specific physiognomic aspects of the sitter – and the portrait as expressive of the sitter's personal qualities. While this tension is constantly lurking between the lines, it is not always directly addressed. I therefore ask of the reader to keep this in mind, so that the final discussion in Chapter 4 will present as the necessary end point it is.

Existing Research and Method

It is beyond question that the topic of this thesis is situated within a discourse on optics and vision, which has been central to art in the Western tradition at least since the Renaissance. Merleau-Ponty, Panofsky, Lacan and Sartre are examples of nineteenth century thinkers who have problematized the importance of the "gaze" in Western society and culture – especially in relation to problems concerning knowledge and politics. So much so, that ocular metaphors are even ingrained in most European languages, as Martin Jay has skillfully demonstrated (1994: 1). While his examples are taken from French and English, this is true for Norwegian as well; especially regarding the use of light-metaphors in relation to knowledge and thinking ("opplyse," "kaste lys på," "reflektere," "være klarsynt," etc.). Clearly, AFR systems are produced on the basis of the idea that visual documentation is an especially trustworthy form of evidence and can thus be read into a long history of technologies serving as prostheses to enhance the human eye. Even modernity itself, as Jay has put it elsewhere, is normally considered "resolutely ocularcentric" (1988: 3).

The discussions in this thesis are also clearly related to what Jonathan Crary has described as "the problem of the observer," informing our very ways of seeing. This, he argues, constitutes "the field on which vision in history can be said to materialize, to become itself visible" (1991: 5). This renders the topic of surveillance, and AFR more specifically, especially well-suited for analysis from an art historical perspective. It is indeed a topic of wide scholarly debate both here and within the broader field of visual studies and the bourgeoning discipline called surveillance studies. The fact that AFR technologies are essentially systems of automated *image interpretation* has nevertheless not been much discussed. Here, Lila Lee-Morrison's doctoral thesis from Lund University stands out. As she also points out, "there has been little scholarship from within the cultural studies and the humanities more broadly on the ways in

which AFR actually performs recognition and how this may constitute and enculture a mode of perception" (2019: 27). Instead, the technical processes have been widely left to the field of computer science, which – as this thesis will make clear – falls short when it comes to grasping the epistemic and political dimensions of their own products. Like Lee-Morrison, I also focus on the performative aspects of AFR, but through a very different set of artistic and technological examples. Her thesis has nevertheless been of much help to me in setting an example for how AFR might be approached from an aesthetically oriented point of view. I also have Lee-Morrison to thank for the idea of comparing AFR's mode of perception to the Wittgensteinian concept of family resemblance, as discussed in Chapter 2. This being said, the field is still at an underdeveloped stage, leaving much to be wanted. This last point will be further developed in Chapter 1, where I discuss how several scholars within the various visual studies are calling for an increased focus on the networked nature of digital photography.

Besides the general stance that the relationship between photographs and computers in AFR constitutes a complicated assemblage of operations and that its production can be understood in light of tendencies associated with the idea of modernism, this thesis has no specific theoretical framework as its starting point. This is a conscious choice on my part, as I am primarily interested in discussing the topic as it is presented by the artworks, rather than reading them into a pre-existing framework. Too often, art is being treated as mere illustrations of ideas rather than as proper epistemic statements themselves. I will not contribute to this. Of course, the idea of art as such is up for debate: the so-called epistemization of art is indeed a hot topic within the contemporary art scene.[1] This is not to say that I am of the opinion that all art is – or should be – concerned with neither knowledge production nor politics. What I *am* arguing, however, is that the projects presented in *this thesis* are, and should therefore be acknowledged as such.

As for the bibliography, the reader will probably notice how it is comprised of a mix of literature from various fields, art history and other visual studies being dominant. In Chapter 4 I have also made active use of work from within surveillance studies. Furthermore, I am cautious to point out that I am *not* a computer scientist, and my understanding of how computer systems work does probably not meet what is considered academic standards within the computer sciences themselves. When discussing these particular issues, I have been reliant on introductory books as well as blogs aimed at explaining computer science to a general audience.

[1] For an illuminating discussion on the epistemization, or "knowledgization," of art, see Tom Holert (2020), *Knowledge Beside Itself: Contemporary Art's Epistemic Politics* (Berlin: Sternberg Press).

However, my goal is not to deconstruct or propose better technical methods for building and using AFR technology, but rather to demonstrate how such algorithms are based on some questionable assumptions regarding both the truth-value of photography and the possibilities of denoting an image's subject matter by a restricted set of words. In this respect, I hope that the reader will find my presentation of the technicalities to be sufficient.

Important Terms

Before moving on to the first chapter, I round off the introduction by making clear what I mean by some terms which are central to the discussion of this thesis. Firstly, it should be clear that when I speak of photography here, I refer to the medium in the broadest sense possible, including everything from daguerreotypes to stills from moving digital image flows. If referring to a specific mode or form of photography, this will be noted in the text. Secondly, as the readers of this thesis may not be schooled within the computer sciences, the term *algorithm* should be explained. Within computer science, this describes *any effective procedure that reduces the solution of a problem to a predetermined sequence of actions* (Nowviskie 2014: 1). There are several uses for algorithms in software, of which the most common are to perform calculations, conduct automated reasoning, and process data. Algorithms can also, amongst other things, be implemented in mathematical models, mechanical devices or electrical circuitry. As one dictionary puts it, the term algorithm in common usage typically references "a deterministic algorithm, formally defined as a finite and generalizable sequence of instructions, rules, or linear steps designed to guarantee that the agent performing the sequence will reach a particular, predefined goal or establish incontrovertibly that the goal is unreachable" (Nowviskie 2014: 1). In this, they are different from *heuristics*, which are processes of trial and error. There exists, however, *nondeterministic algorithms* which are designed to solve harder problems by finding the best solution available within a given set of restraints (Nowviskie 2014: 1). The algorithm discussed in in Chapter 2 is an example of the latter.

The last term which should be explained is *data*, as its colloquial use may obscure the technical meanings. According to one dictionary, data come in three kinds: the first is "data that refers to something outside of itself – encoding or representing changes outside of the computer." Next, there are data *as data*, simply meaning anything handled by a computer. Lastly, there are data that *work on data*, i.e. a *program* (Fuller 2014: 125). The latter definition refers to how most computer programs are themselves made up of data. In this sense, data is worked *on*, but also *does work*. The consequences of this will become clear throughout the thesis.

Chapter 1

Less Than a Thousand Words: ImageNet Roulette by Trevor Paglen and Kate Crawford

Introduction

Whenever the subject of machine vision in art is discussed, the name Trevor Paglen almost inevitably comes up. This American artist's work seems to be included in almost every single exhibition, talk or text related to this topic. Paglen (b. 1974) is known for his many projects concerning surveillance and state secrecy. Through a highly multi-faceted practice he has addressed the issue of covert global surveillance at least since the early 2000's. Already holding an MFA from the Art Institute of Chicago, Paglen also obtained a PhD in geography from the University of California, Berkeley in 2008. This speaks of the artist's multidisciplinary approach to his work. While centered around photography, Paglen's artistic practice has gradually developed into spanning a wide range of mediums including performance, sculpture and writing – often made in collaboration with teams of differently trained professionals. In the early 2010s, Paglen gained widespread international attention in the artworld with a series of projects which all focused on the material traces of clandestine military operations and finding ways to literally make these processes visible. Amongst these, *The Black Sites* (2006), *The Other Night Sky* (2010 -), and *Limit Telephotography* (2004-12) seem to be the most widely exhibited and praised. Since then, Paglen's work has been exhibited, collected and discussed by a wide range of internationally acclaimed art institutions, rendering any survey of surveillance related art incomplete without mention of his work.[2]

As undertaking a thorough analysis of Paglen's whole oeuvre would be a too comprehensive task in the context of an MA-thesis, I have chosen to discuss those of his projects that specifically addresses the topic of machine vision. This implies emphasizing the part of his work that followed after the performance *Sight Machine* in 2017. From that time on, he seems to have been more or less exclusively focused on instances of machine vision. After becoming an artist fellow at the AI Now Institute[3], he has already produced at least eighteen projects specifically examining this phenomenon, amongst which several concern the analysis

[2] A full timeline of Paglen's life and work until 2017 is available in the monograph Trevor Paglen (2018) by Cornell, Bryan-Wilson and Kholeif (London: Phaidon).
[3] The AI Now Institute at the New York University was founded in 2017 with the goal to produce interdisciplinary research on the social implications of artificial intelligence.

of faces by computer programs.[4] The discussion in this chapter will start with a presentation of the project *ImageNet Roulette* (2019), which was a collaborative work by Trevor Paglen and Kate Crawford, who is a leading scholar on the social and political implications of AI. What they produced was more precisely a software, or "app", embedded in a webpage to which the audience were invited to upload photos of themselves in order to have them categorized by an algorithm trained in the image database ImageNet. I will then go on to discuss the related exhibition *Training Humans*, which opened in Milan shortly after the release of the app in September 2019. All in all, the different parts of the project seem to raise questions of representation and power: what politics underlie the labeling of people into certain categories, and how does the taxonomy in question affect the people whose photographs are unknowingly being utilized by these systems? This brings forth yet another set of questions regarding the relationship between language and image, as the categories are all formulated as single nouns.

Regarding the order of presentation, I have chosen to stick to chronology. This implies presenting the app and how it went viral first, then the essay *Excavating AI*, and lastly the exhibition *Training Humans.* This is both to give the reader a sense of narrative, but also because it will give grounds for understanding the set of prior knowledge that the audience might have brought along with them to the exhibition. As the app gained a lot of international attention, I find it likely that many exhibition visitors would have registered the discussion it generated beforehand. The experience of the two can thus be seen as continuous, the one informing the other.

The Viral App

In September 2019, the hashtag #*ImageNetRoulette* was trending on Twitter and other social media. It referred to the selfie-app *ImageNet Roulette*, which had gone viral. The concept was simple: the audience, or "user", uploads a photograph of a person, and the software detects a face in it, which it labels with one of nearly 3000 person categories from the database ImageNet. When first released, the app appeared as one amongst many popular selfie apps where you can upload your photo and it will let you know "what you would look like as a toddler", "what artwork features your doppelganger" and the like. *ImageNet Roulette*, however, proved to be something else entirely. As more and more people posted tagged images of themselves expressing how the algorithm "saw" them, *#ImageNetRoulette* was gaining lots of attention in

[4] On his own webpage, Paglen har categorized his various projects by topic. I am here relying on the artist's own classifications of his work. URL: https://paglen.studio/category/machine-visions/

social media almost immediately after its publication.[1] This is indeed how I learned of the project myself. And how did I, a 25-year-old bespectacled white woman get labeled? "Eccentric, eccentric person, flake, oddball, geek: a person with an unusual or odd personality." Experimenting, I also uploaded another portrait where I had taken my glasses off and ruffled my hair. The app then tagged me as "caveman, cave man, cave dweller, troglodyte: someone who lives in a cave." In most cases, like mine, the results seemed harmless, even fun, although very banal. But soon it became clear to anyone following the trend how some of the tags were also highly problematic.

A tweet that caught my attention was posted on September 18 by a user named Lil Uzi Hurt (@lostblackboy), writing "no matter what kind of image I upload, ImageNet Roulette, which categorizes people based on an AI that knows 2500 tags, only sees me as Black, Black African, Negroid or Negro" (2019). Along with this text, he posted four photos of himself in various situations. They show a young and well-dressed man in a variety of situations, seemingly hand-picked in order to reflect his multifaceted personality. In all of them, his face is framed by a neon green square along with the tags "black, black person, blackamoor, Negro, Negroid" (Fig. 1). The labels in green show how, what to me look like deeply human photographs full of personality, are perceived very differently by a software that does not see past the color of this man's skin. Another example is told in a New York Times article, where the app indeed *does* come up with labels denoting personal character. The labels, however, are not exactly flattering in this case either: Mr. Kima, a 24-year-old African-American, experienced that "when he uploaded his own smiling photo, the site tagged him as a 'wrongdoer' and 'offender'" (Metz 2019). This demonstrates racist stereotyping which appears more problematic than ever at a time where police violence and racial profiling by police is an especially pregnant topic in light of the Black Lives Matter movement.

A wide range of similar posts and articles demonstrate how the algorithm employed by *ImageNet Roulette* is clearly biased in several ways, of which racist and misogynist tendencies are most prominently featured. The Guardian journalist Julia Carrie Wong gives words to the experience of being stigmatized by the software (Fig. 2):

> (…) after a day of watching my fellow journalists upload their ImageNet Roulette selfies to Twitter with varying degrees of humor and chagrin about their labels ("weatherman", "widower", "pilot", "adult male"), I decided to give it a whirl. That most of my fellow tech reporters are white didn't strike me as relevant until later. I don't know exactly what I was expecting the machine to tell me about myself, but I wasn't expecting what I got: a new version of my official Guardian headshot, labeled in neon green print: "gook, slant-eye". Below the photo, my label was helpfully

> defined as "a disparaging term for an Asian person (especially for North Vietnamese soldiers in the Vietnam War)" (Wong 2019).

Together, these comments resulted in public shock and outcry, making the app gain ever more attention. At one point, it was generating as much as 100 000 labels an hour, according to the New York Times (Metz 2019). At this point, only a few days after its release, many were probably asking themselves how yet another selfie app with clear biases could come to be released without these problems being discovered beforehand. For example, an 2018 app by Google Arts Project made to let people find their doppelgangers in museum collections had been widely criticized for including a too narrow variety of Asian art so that Asians and Asian Americans were matched with pictures that looked nothing like them (Goggin 2019). Already three years previously, Google had also been criticized because one of its facial recognition programs tagged people of African descent as gorillas (Simonite 2018). While many were probably inclined to attack *ImageNet Roulette* as a continuation of these problematics, it soon became clear that the revelation of inherent biases was actually an intended part of the project. According to Paglen, the bigger point that he and Crawford were trying to make was "how dangerous it is for machine learning systems to be in the business of 'classifying' humans and how easily those efforts can – and do – go horribly wrong" (Rea 2019). To spur criticism of AFR technology was in other words exactly the intended goal behind the roulette.

The point seemed to get across to a lot of people. In fact, *ImageNet Roulette* gained so much attention that it after only a few days resulted in a publication by ImageNet, the database which it critiqued, stating that 1,593 out of the database's 2,832 person categories were now deemed as "unsafe", and that over 600 000 images had been removed from the database (Yang et al. 2019). Exactly what this means is left unclear, but we are led to believe that the "unsafe" categories will not be used for training algorithms in the future. Although the statement does not actually include any references to neither Crawford nor Paglen, the timing seems too close to the app's release to be coincidental. In this respect, *ImageNet Roulette* appears to be a great success if understood as a form of activism, a view that Paglen and Crawford seemed to share. Not long after, the two announced that they would shut down the webpage from September 28 because the project, as they saw it, had made its point: "it has inspired a long-overdue public conversation about the politics of training data, and we hope it acts as a call to action for the AI community to contend with the potential harms of classifying people" (Artforum 2019).

Thus, less than a month after its release, the webpage with *ImageNet Roulette* was already gone. If you tried to visit it now, you would be redirected to another webpage featuring an essay contextualizing the project. As the actual software was made by Paglen and Crawford

themselves and only *trained* in ImageNet, it was not obvious how the biases it revealed actually related to the database itself. In my view, the text replacing it was a welcome next step as it answered the kind of questions that the roulette alone had left open. In the following, I will present the main arguments of this essay, *Excavating AI*, before moving on to present the exhibition *Training Humans.*

Figure 1: *Screenshot of tweet by the user @lostblackboy posted on 18.09.2019. Retrieved on January 26, 2021.URL: https://twitter.com/lostblackboy/status/1174112872638689281*

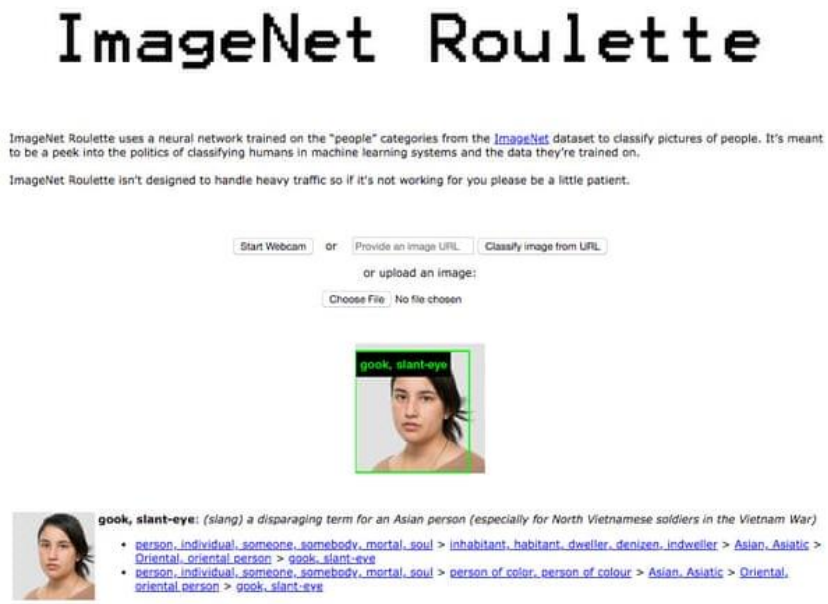


Figure 2: *Screenshot of the* ImageNet Roulette *webpage posted by the Guardian journalist Julia Carrie Wong. Retrieved on 26.01.2021.URL: https://www.theguardian.com/technology/2019/sep/17/imagenet-roulette-asian-racist-slur-selfie.*

An Archeology of Datasets

Excavating AI is introduced with a short summary of how the earliest attempts of developing automated image recognition systems in the 1960s proved to be a much more complicated task than first expected. Ever since, the problem has, according to the authors, been approached in purely technical terms, seeking to axiomatize the meaning of images. The same goes for facial recognition programs, as they are essentially a specialized form of image recognition. Asking whether the whole idea of approaching AFR as a technical problem is misinformed, Crawford and Paglen position themselves in opposition to what they portray as the norm within the field of AI. As an alternative, they want to "explore why the automated interpretation of images is an inherently social and political project, rather than a purely technical one" (Crawford & Paglen 2019). The essay can thus be understood not only as an attempt to discuss the particular database ImageNet, but also to shed light on widespread attitudes informing the AI community more broadly.

Based on a method they refer to as an *archeology of datasets*, the authors pose some overarching questions: first, they will discuss what work images *do* in AI systems, and secondly what computers are meant to recognize in an image – as well as what is *mis*recognized, or even completely invisible (Crawford & Paglen 2019). Although the authors put a special focus on ImageNet, which they refer to as "the most iconic training set of all," the reader is also introduced to several other datasets meant for training facial recognition algorithms. In this way, Paglen and Crawford shed light on some questionable assumptions which seem to be inform various taxonomies employed across the field of AFR, giving credence to the idea that there is an underlying problem epistemological in nature. The following is a summary of how the authors explain automated facial recognition in *Excavating AI*. I dedicate this a considerable amount of space because it also serves the goal of this chapter well: to analyze ImageNet Roulette while also to introducing the building blocks of AFR systems more generally.

Training sets are the foundation on which machine learning[5] systems are built. If you want to build any kind of AI system, you need data on which to train your algorithm. If the task you want to teach is object or facial recognition, you will need vast amounts of carefully labeled images which are sorted into categories. On a basic level, a training image dataset (TID) can

[5] Machine learning seems to be the preferred term used by practitioners within the field, instead of "AI", which is much less specific. For more on this, see for example, Mattheo Pasquinelli (2019), "How a Machine Learns and Fails" in *Spheres* No. 5. Machine learning is also the term which the authors actually use in this text even though they use AI both in the title as well as in various interviews discussing the project. I therefore come to wonder if this is a strategic phrasing chosen for marketing purposes, as AI has different connotations and therefore might incite different responses with the general public.

thus be understood as a digital archive of images before the training even begins. When the algorithm is trained in the TID, what it does is to convey a statistical survey of the many image-label combinations before creating a model of what distinguishes one class of items from another by analyzing the data contained in the images. Paglen and Crawford illustrate this process with an example where the task is to make a system for recognizing the difference between apples and oranges:

> A developer has to collect, label, and train a neural network on thousands of labeled images of apples and oranges. On the software side, the algorithms conduct a statistical survey of the images, and develop a model to recognize the difference between the two "classes." If all goes according to plan, the trained model will be able to distinguish the difference between images of apples and oranges that it has never encountered before (2019).

The actual accumulation and labeling of images, then, is done manually by people, while what the algorithm does is simply to conduct a statistical analysis of this data and compare new images with its previous calculations in order to match it to a category. When it comes to categorizing images of *faces*, the process is exactly the same even though the subject matter is different. Whatever biases are learned are hence direct results of the categorization processes done while creating the TID. This is precisely why Paglen and Crawford find it so important to bring attention to the training datasets themselves.

There are several reasons as to why ImageNet is of particular interest to this discussion. While originally presented as a research paper, ImageNet was later constructed and has grown to constitute a dataset of "extraordinary scope and ambition" (Crawford & Paglen 2019). Setting out in 2009 with the ambition to "map out the entire world of objects", it had after ten years scraped over 17 million images from the internet (Crawford & Paglen 2019). This was achieved by using what in *Excavating AI* is referred to as "an army of piecemeal workers" hired from Amazon Mechanical Turk.[6] According to Crawford and Paglen, these workers were sorting an average of 50 images per minute according to the more than 20 000 categories available. In the following decade, ImageNet has been extremely important in the field of image recognition and research within computer vision more generally. For example, the database has been the basis for an annual competition where different teams of programmers, or "labs", test

[6] Amazon Mechanical Turk is an online crowdsourcing platform through which companies can hire people to do on-demand manual work that computers are not currently able to do. These workers are often referred to as "Turkers" and are known to be paid extremely low wages for their work. The name Mechanical Turk is derived from an automated chess player constructed by Wolfgang von Kempelen in 1770. It toured Europe, playing very strong games against many people, including Napoleon Bonaparte. The machine was later revealed to be a hoax: the mechanical "turk" was in fact operated by a human hiding within it.

their algorithms by pitting them against a given subset to see who can achieve the lowest error rate.[7] It was considered a turning point in AI history when a team from Toronto used a *convolutional neural network* (CNN)[8] to win the competition's top prize in 2012. This led to an extreme increase in accuracy across the field. By 2017 (the final year of the competition) the record was 97,3%, which I take this to mean that the accuracy rates are very good (and it is surely even higher at this point). This means that the problems of bias do not lie in the algorithms inability to correctly follow the given set of instructions. Therefore, the issue must rather lie at the point of determining *what is deemed an accurate interpretation*.

In order to understand the taxonomy of ImageNet, one first has to know about the word classification database WordNet, on which its semantic structure is based. Wordnet was developed in the 1980s (also at Princeton), and is organized according to a nested structure of so-called synsets (cognitive synonyms). Each synset represents a distinct concept where synonyms are grouped together. The relationship between these is explained by Crawford and Paglen as follows: "synsets are then organized into a nested hierarchy, going from general concepts to more specific ones. For example, the concept 'chair' is nested as artifact > furnishing > furniture > seat > chair" (2019). As opposed to WordNet, which attempts to organize the *entire English language*, the synsets in ImageNet are restricted to nouns – supposedly because this class of words is thought of as encompassing the entirety of what pictures can represent.

Regarding hierarchy, there are nine top-level categories under which all the others are divided: plant, geologic formation, natural object, sport, artifact, fungus, person, animal, and miscellaneous. While it is primarily the person category which is the object of Paglen and Crawford's discussion, it should be pointed out that *any* taxonomy or other system of classification is always fueled by politics. I quote one paragraph of the essay here in full, as I believe it captures this problem very well while also pointing towards some historical precursors to this practice:

> The category "human body" falls under the branch Natural Object > Body > Human Body. Its subcategories include "male body"; "person"; "juvenile body"; "adult body"; and "female body." The "adult body" category contains the subclasses "adult female body" and "adult male body." We find an implicit assumption here: only "male" and "female" bodies are "natural." There is an ImageNet category for the term "Hermaphrodite" that is bizarrely (and offensively) situated within the branch Person > Sensualist > Bisexual > alongside the categories "Pseudohermaphrodite"

[7] In ImageNet's defense, it should be noted that the person categories have, as far as I know, never been included in this competition.
[8] CNNs are explained in more detail in Chapter 2.

> and "Switch Hitter." The ImageNet classification hierarchy recalls the old Library of Congress classification of LGBTQ-themed books under the category "Abnormal Sexual Relations, Including Sexual Crimes," which the American Library Association's Task Force on Gay Liberation finally convinced the Library of Congress to change in 1972 after a sustained campaign (2019).

It is apparent that already on the level of taxonomy, politics are embedded in the process of deciding what categories to include and where the various synsets are deemed to belong. That "bisexual" and other categories have been included under the branch "sensualist" tells of a certain moralism ingrained in the process of creating the database itself; making judgements on what sexualities or lifestyles are considered "sensualist" as opposed to other ways of being/living. This renders the taxonomy offensive already before the point of connecting the categories to images. The fact that the creators of ImageNet apparently believe personal properties such as the above to be potent visual identifiers in photographs just makes this even more problematic, on the verge of absurdity. As anyone within the fields of art history, media theory and the like will know, the authors point out, images are "slippery things" whose full meanings are seldom easy to describe with words – let alone with a single noun. Yet, as ImageNet Roulette demonstrates, the assumption that this can be done is exactly what TIDs are based on.

At the time when *Excavating AI* was written, there were 2,833 subcategories under the top-level category "Person," labeling images of people according to attributes such as race, nationality, profession, economic status, behavior, character and morality. Although the assumption that any of these can be identified visually is problematic, it is with the latter three categories that the system seems to take a really dark turn. Paglen and Crawford summarize:

> There are categories for Bad Person, Call Girl, Drug Addict, Closet Queen, Convict, Crazy, Failure, Flop, Fucker, Hypocrite, Jezebel, Kleptomaniac, Loser, Melancholic, Nonperson, Pervert, Prima Donna, Schizophrenic, Second-Rater, Spinster, Streetwalker, Stud, Tosser, Unskilled Person, Wanton, Waverer, and Wimp. There are many racist slurs and misogynistic terms (2019).

The examples cited above are only some amongst many possible that could have been mentioned. Still, even when looking at only these, it becomes clear how this very important TID is based on a taxonomy which is far from neutral, exhibiting gendered, racialized, ableist and ageist biases – just to name some of the most obvious problems. What gives someone the right to decide what a "bad person," a "loser" or a "failure" looks like? To make matters worse, this kind of classification system has become more and more common in recent years in small and big AI companies alike. While perhaps the most important one, ImageNet is far from being the only database of its kind. In the essay, Crawford and Paglen also introduce the reader to a

couple of others; three of which are described at some length: JAFFE, UTKFace, and IBM's "Diversity in Faces."

JAFFE stands for Japanese Female Facial Expression and is a dataset created in 1998 to train algorithms to recognize facial expressions. It consists of photographs of ten Japanese women with seven different facial expressions. The first assumption being put to questioned here, is the idea that concepts like emotion can be applied to people's faces at all. The second is why these researchers think that there are only six emotions plus a neutral state? Third, there is the obvious problem that even if the first two assumptions are verified, one is still left with the fact that the emotions displayed in the pictures are *acted out* by the women and therefore do not express their actual feelings or state of mind. These problems clearly reduce the credibility of this dataset, but problems aside, JAFFE appears relatively harmless compared to the other TIDs discussed in *Excavating AI*. Having been made before it became possible to scrape millions of images from the internet – and also before the availability of cheap online labor through crowdsourcing platforms – the scale and ambition of JAFFE appears modest.

When it comes to TIDs produced in the age of Internet 2.0, on the other hand, the case is very different. Crawford and Paglen note how, "as training sets grew in scale and scope, so did the complexities, ideologies, semiologies, and politics from which they are constituted" (2019). UTKFace was made by a group of researchers at the University of Tennessee at Knoxville (hence the acronym) and consists of about 20,000 images of faces with annotations for gender, age and race. According to the authors, it can be used for tasks like automated facial detection as well as age estimation and age progression. What they find most troubling about this dataset is how these categories are limited to a very small number of subcategories: gender is presented as a binary choice between male and female, while age is limited from 0 to 116. While these assumptions can certainly all be questioned, the most troubling category, from my perspective, is nevertheless the one that places people into one of five racial classes: White, Black, Asian, Indian and "Other".

It is unclear how this simplistic idea of race was determined. The taxonomy is nevertheless clearly problematic, recalling some of the racial categorization systems of the nineteenth and twentieth century. As the authors point out, UTKFace seems to parallel the so-called *Book of Life* used by the Apartheid regime in South-Africa from around 1970. With these "identity passbooks," the country's entire population was categorized as either Black, White, Colored or Indian, and how people were identified had far-reaching consequences for their rights and freedoms. As Crawford and Paglen point to, the dataset can also be compared to the nineteenth century context of social Darwinism and imperialism in the nineteenth century, when

pseudo-sciences like physiognomy, phrenology and eugenics were at their prime. As I will discuss in depth in Chapter 3, all of these disciplines sought to classify humans based on physical attributes under the pretext of determining risk of deviant or criminal behavior. What they in practice actually studied, however, was "deviance from bourgeois ideals" (Crawford & Paglen 2019).

The database *Diversity in Faces* is no better than the ones discussed above. It was released by IBM in 2019 as a response to criticism of the company's existing facial recognition system regarding problems with recognizing people with dark skin tones. The new system was meant to be more representative, constructed as a "computationally practical basis for ensuring fairness and accuracy in face recognition" (Crawford & Paglen 2019). In line with what appears to be the norm when constructing TIDs, the company hoarded hundreds of thousands of images of unsuspecting people from the internet, especially from Flickr. What is special about this particular dataset, however, is the precise manner in which the developers decided to sort the images afterwards. Unsure of whether including age, gender and skin color in the top categories was sufficient to ensure "fairness and accuracy" – qualities which are themselves not easily defined – IBM decided that *even more* classification seemed like a good strategy. Thus, understanding the problem in purely quantitative terms, IBM moved into what Paglen and Crawford refer to as "truly strange territory": in addition to the other three categories, they decided to add *skull shape* and *facial symmetry* to the list.

By taking a mathematical approach in their process of cognitive bias mitigation, IBM's Diversity in Faces recalls yet another nineteenth century pseudoscientific method. In this case, the taxonomy of the dataset can be seen as a direct parallel to craniometry, which was used until the early 1900s as a way of predicting intelligence based on skull shape and weight (Crawford & Paglen 2019). Reflecting on this, Paglen and Crawford state as follows:

> Ultimately, beyond these deep methodological concerns, the concept and political history of diversity is being drained of its meaning and left to refer merely to expanded biological phenotyping. Diversity in this context just means a wider range of skull shapes and facial symmetries. For computer vision researchers, this may seem like a "mathematization of fairness" but it simply serves to improve the efficiency of surveillance systems (2019).

What the authors point to here is how programmers involved with facial recognition handle the concept of "fairness" without ever asking whether the problems they face might rather lie in the structure of the datasets or in the questions they are designed to answer. To the extent that they acknowledge the existence of any problems, they tend to use them as an excuse to harvest *even more data.* The lesson I take away from *Excavating AI* is thus that the biases in *ImageNet*

Roulette cannot be depreciated simply as "glitches" or "bad programming." Instead, they point to deep systemic issues.

While the TIDs presented above each have their individual quirks, they are all based on the shared assumption that personal qualities can somehow be identified visually in photographic portraits. In several cases, this results in taxonomies that share the racist assumptions which were preeminent in the age of physiognomy, a matter which I will discuss in depth in Chapter 3. While the technical issues of creating a taxonomy are certainly interesting, much deeper issues also arise in the discussion, such as what agendas lie behind the creation of AFR technology in the first place, and what gives these tech-companies the right to be the judges of what is considered "fair." Lastly, copyright and privacy concerns are also clearly relevant in that the photos employed are scraped from the internet without neither the consent nor knowledge of the people who are depicted.

Training Humans

If you thought that the story of *ImageNet Roulette* was over by now, you were wrong. As the many articles featuring the project will let you know, the app was created in concert with an exhibition opening its doors at the Fondazione Prada in Milan about the same time in September 2019. The exhibition was said to include some of the same images as those presented in *Excavating AI*,[9] which created expectations for it to build on the lessons from the essay. Having received very positive reviews – both Digicult (2019) and Mousse Magazine (2019) described it as a *landmark exhibition* – my expectations of *Training Humans* were naturally high. I was also triggered by Fondazione Prada's website, which brazenly announced *Training Humans* as the first ever major photography exhibition devoted to training images. Furthermore, it was promised to "reveal the evolution of training image sets from the 1960s to today" by exploring "how humans are represented, interpreted and codified through training datasets, and how technological systems harvest, label and use this material" (2019).

The Gallery's promotional text posed two questions, around which the exhibition was supposedly centered: "where are the boundaries between science, history, politics, prejudice and ideology in artificial intelligence? And who has the power to build and benefit from these systems?" (Fondazione Prada 2019). While these questions were discussed to some extent in *Excavating AI*, the exhibition was promoted in a way that gave the impression that one might expect to learn even more about the technicalities involved here. A statement by Crawford

[9] This is, amongst other places, pointed out in the list of acknowledgements in *Excavating AI*.

substantiated this perception: “what we hope is that ‘Training Humans’ gives us at least a moment to start to look back at these systems, and understand, in a more forensic way, how they see and categorize us” (Fondazione Prada 2019). To what extent could such claims be substantiated? In the following, I will discuss this question, starting with a presentation of my own experience of the exhibition.

In February 2020, I found myself at the Osservatorio; Fondazione Prada’s art gallery located in the exclusive Milanese indoor shopping street Galleria Vittoria Emanuele II. To get there, the gallery visitor first had to go inside the fashion brand’s flagship store and take the well-hidden elevator to the fifth floor. Passing the reception, you would find yourself standing inside an interior reminiscent of a spacious penthouse apartment. The walls were painted in a calming shade of blue, the floors covered in a stately fishbone parquet, and large windows along the whole left side of the gallery let fresh daylight fill the space. The immediate impression of the gallery space was more inviting than its rather exclusive (in the true sense of the word) surroundings, but this feeling was countered by an eerie soundtrack. Repeating a series of voices reciting what sounded like randomly assembled sentences in a variety of English accents, this hidden source of sound sharpened one’s senses. Looking inwards, the first thing I noticed was a large black and white print of a highly pixelated fingerprint dangling from the ceiling in the middle of the room. Behind it, to the right, was a long wall containing large amounts of photographs arranged in clusters, as well as several flickering TV-screens – some mounted on the wall and others on the floor. Somewhat overwhelmed by the sheer number of images presented here, the first impression of the room was rather confusing even though the many images did not physically take up much space.

Uncertain of what dataset to approach first, I turned and looked at the welcoming text on the wall behind me. It started by ascertaining that the relationship between humans, images and image-making technologies has changed dramatically during the last decade with AI becoming ever more present in everyday lives. Furthermore, besides a list of points already discussed in *Excavating AI*, the text also explained the general layout of the exhibition: it was organized chronologically with the first (fifth) floor presenting the earliest TIDs from the 1960s onwards, while the newest datasets would be presented on the second (sixth) floor. The text then concluded with the following, rather leading statement, promising a certain build-up of intensity as the visitor would move along:

> As the classification of humans by AI systems have become more and more invasive and complex, their biases and politics have become apparent. Within computer vision and AI systems, forms of measurement turn into moral judgements. Our images now look back at us. And we won’t always like what – or how – they see.

My interest piqued, I continued my tour, curious to uncover how Paglen and Crawford had solved the problem of representing this kind of interactivity on behalf of images.

In the middle of the room, behind the fingerprint, I encountered a large, vertical flat-screen TV flickering between short clips of people wearing a set of wires while walking on a mat inside a lab. A study of people's gate, perhaps? Behind it to the left was another TV, only a much older model. Looking at it, I recognized the footage from the notorious Blakemore and Cooper 1970 experiment where the two scientists let kittens grow up in visually limited surroundings in order to study the impact of visual stimuli on the development of the primary visual cortex. In only a matter of short time, it was clear to me that the curators intended to bring forth a sense of being monitored, and it worked. When I now turned towards the wall featuring all the photographs, the feeling of being watched was palpable. Here, there were two sets of images separated by a small TV-screen flickering between a series of high-contrast black and white portraits of faces in various angles and lighting conditions. This peculiar, square screen further manifested the eerie atmosphere in the gallery space. To its right was a group of photographs organized in straight rows (Fig. 3). These were close-up portraits of faces framed from the shoulders up, all in front of the same grayish background. The photographs depicted men and women of various ethnicities, all printed in the same "a little smaller than real-life" size. Besides the framing, other common features were the subjects' somewhat strained or unnatural expressions, as well as 1990s styling on hair and make-up. These portraits were clearly taken for research purposes.

Contrasting these heavily stylized photographs, the left-hand images were a set of mugshots displaying prisoners in orange jumpsuits staring out at you with exhausted expressions (Fig. 4). For anyone acquainted with the poor privacy rights of people incarcerated in the US, it was easy to imagine how such pictures must be easy prey for anyone looking for a cheap collection of pre-labeled images. Upon further inspection, it became clear to me that each row consisted of photos of the same person, but at different moments; an observation which was affirmed by the title *Multiple Encounter Dataset*. This made me ponder whether the dataset had been used for studying changes in appearance over time. In any case, looking at this dataset certainly raised many questions and also served to connect the AFR industry to other controversial businesses. Before continuing to the second gallery floor, it had thus already been made clear through looking at the different documentations that machine vision is not limited to AFR, but that such technologies are ingrained in a multilayered matrix of businesses centered

around the surveillance and analysis of our expressions, movements, and – as with the kittens – even the development of vision itself.

When reaching the top of the stairs leading to the second gallery floor, it was immediately apparent that the chronology of the exhibition had reached the era of Internet 2.0, and consequently the time of mass-harvesting of online images (Fig. 5). Here, the photographs of people were organized in rows with no space between them. Each image was also much smaller than the ones encountered downstairs. Moreover, the TV screens on this floor were placed amidst the assemblage of photographs, blending in instead of separating them. The pace at which the screens flickered between images had also been sped up considerably. It was as if the effect of the internet on the culture of images was symbolized through the literal reduction of each photo, and that the inclusion and juxtaposition of so many at once was meant to be read as a direct material translation of the exponentially growing number of images uploaded online. The rapid frequency at which the screens were shifting between images also suggested some very important differences between machine vision as opposed to how humans look at images: Not only are computers able to do this at a much, much faster speed, but the manner in which this whole practice is enacted is also completely different. Where humans *look* or *observe*, the machinic vision represented here rather appeared as a process of *scanning.*

In the middle of the room was a large vitrine table. When walking past it, towards the wall, I recognized the Japanese women from the JAFFE dataset on the one side, and UTKFace on the other, both of which were discussed in *Excavating AI*. As there did not seem to be anything new about this particular representation of the datasets (besides being printed out), however, I continued towards the long wall of photos. Standing close, one could immediately recognize a series of celebrities in the photographs, such as Michael Jackson, Tupac Shakur and Angelina Jolie. Upon reflection, the layout was actually reminiscent of a Google search; the rows of photos representing the search results. I also noticed that there were especially many photographs of Angelina Jolie, which worked as a reminder of how celebrities are particularly vulnerable to facial recognition technologies, as photographs of them are spread widely across the globe.

Figure 3: FERET dataset, *National Institute of Standards, 1993-96. As presented in* Training Humans. *Photo: Tuva Mossin.*

Figure 4: Multiple Encounter Dataset II (MEDS-II), *National Institute of Standards, 2011. As presented in* Training Humans. *Photo: Tuva Mossin.*

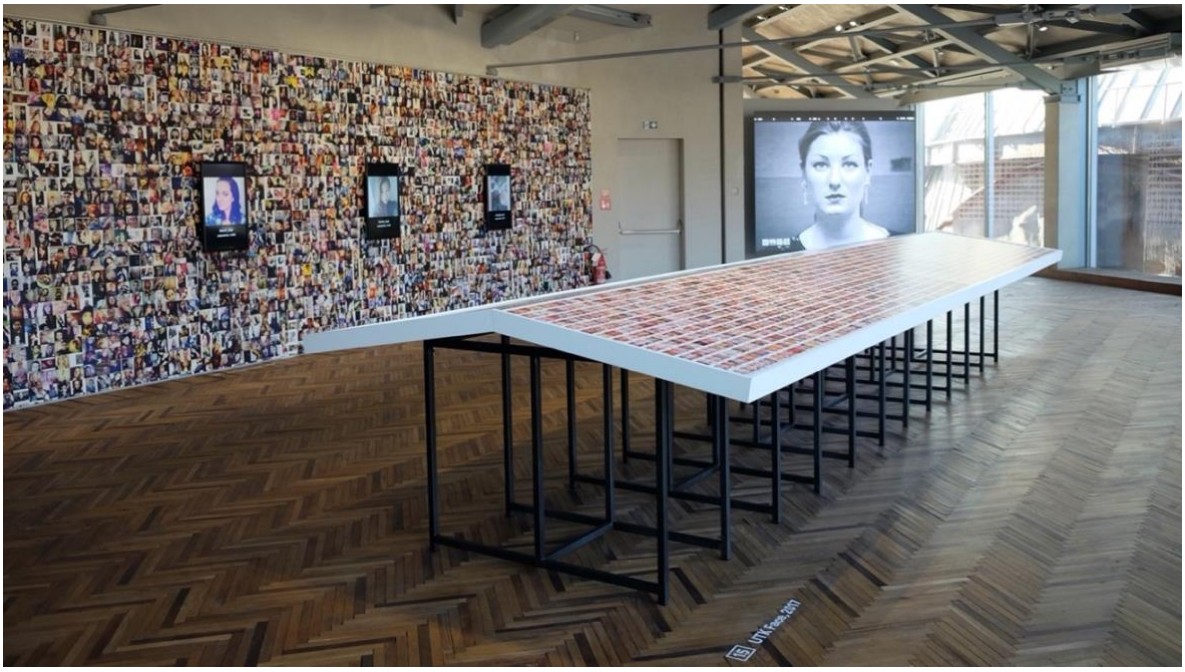


Figure 5: *The view greeting visitors when entering the second gallery floor. Photo: Tuva Mossin.*

Despite the fact that there were many more images on this floor than the one below, the number of datasets was about the same. However, the most considerable share of space was devoted to the representation of ImageNet. Continuing around a partitioning wall covered by yet another large screen showing yet another series of faces, I now found myself standing in a space whose walls were completely covered from top to bottom by photographs in a format reminiscent of Instagram-posts: square-cut photos on white background with captions in black underneath. The captions were categories along with definitions, similar to those in *ImageNet Roulette*. In total, these images covered four walls, and visitors had to turn several corners to see them all. In fact, this presentation of ImageNet included so many photographs that looking at every single one of them would have taken several hours, if not days. As having a good look at every single image was therefore unfeasible, I rather let my eyes glide across the unity of them all, my gaze fixing randomly. The following examples are just a few amongst the many photo-label-combinations I noticed: Barack Obama labeled as an *amoralist* (Fig. 6); a lightly dressed Jennifer Aniston labeled as a *frotteur* (Fig. 7); two children labeled as *bastard* (Fig. 8)*;* and a middle-aged man with a slightly worn face labeled as a *pederast* (Fig. 9).

These observations, along with other possible examples that are too many to be recounted, raised a series of questions. One, which has already been discussed, is the curious fact that the category titles were all limited to nouns, as if this embraces all that which can be expressed in an image. The second question regards the application of the labels themselves:

how does someone decide what an amoralist, a frotteur, a bastard or a pederast looks like? The difficulty of matching single terms with photos of people reminds me of a scene in Maggie Nelson's novel *The Argonauts*. Here, the author vents her frustration as a friend chooses the term *heteronormative* to describe a photo of her dressed-up family printed on a mug:

> But what about it is the essence of heteronormativity? That my mother made a mug on a boojie service like Snapfish? That we're clearly participating, or acquiescing into participating, in a long tradition of families being photographed at holiday time in their holiday best? That my mother made me the mug, in part to indicate that she recognizes and accepts my tribe as family? What about my pregnancy — is that inherently heteronormative? Or is the presumed opposition of queerness and procreation (or, to put a finer edge on it, maternity) more a reactionary embrace of how things have shaken down for queers than the mark of some ontological truth? As more queers have kids, will the presumed opposition simply weather away? Will you miss it? (…) What about the fact that Harry is neither male nor female? (2015: 13).

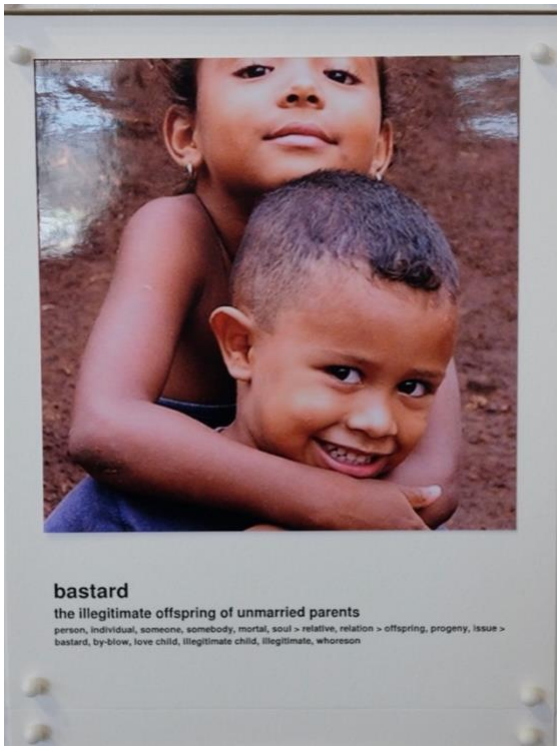
Nelson goes on in more detail, but I believe this excerpt catches the difficulty of denoting the meaning of images with words – especially terms as complicated as this one. Although "heteronormative" differs from the terms cited above in being an adjective, I would claim that determining the essence of what makes someone look like a "bastard" would be just as tricky. A more disturbing question, however, is what use this labeling can possibly have. What are the purposes of labeling images of people with terms such as these? No matter what the answers to these questions are, it was obvious from having encountered ImageNet Roulette that the project as a whole documented some curious visual practices.

Disturbed and very overwhelmed, I turned the last corner of the exhibition and encountered two separate flat-screen TVs standing alone in the middle of this innermost room. This time, I found myself facing real-time video footage of myself (Figs. 10 & 11). At first, I was quite startled by this fact, but fatigued after having studied so very many photographs, this sudden interactivity was a welcome relief. Around my face were the familiar neon green squares known from *ImageNet Roulette*, and as I moved closer to the nearest screen, I could see that I had been labeled as a *psycholinguist*. Thoughtful of this seemingly random categorization, I turned to the other screen, and found the software to be guessing my age, gender and emotion: age 25-32 (0.73); gender female (0.99); emotion fear (1.00). The decimals signified the program's own estimation of the accuracy of its guess. Apparently, while being only 73 percent certain of my age (which was in fact correct) and 99 percent certain of my gender (also correct), the program was curiously one hundred percent certain that my emotional state was that of fear.

Responding to the absurd situation I found myself in, I started laughing, and the label changed accordingly to *happy*. The exhibition was thus concluded by taking the visitor full

circle back to *ImageNet Roulette*, only this time in the form of a video software. Besides the questions already discussed, I had gone about the exhibition wondering whether *my* photo was part of a database like this. At one point, I actually started to look around to see if any surveillance cameras were installed in the gallery. And just as the paranoia of being watched started to wear off, I turned the last corner and found myself facing real-time footage of myself. Thus, by reenacting exactly the kind of non-consensual undermining of privacy that the exhibition critiques, the narrative of *Training Humans* ended in a crescendo confirming the fear that had been nagging on my subconsciousness during the whole visit: *yes, this could be you*.

Figure 6 & 7: *Barack Obama labeled as an amoralist, and Jennifer Aniston labeled as a frotteur. From the dataset ImageNet, as presented* in Training Humans. *Photo: Tuva Mossin.*



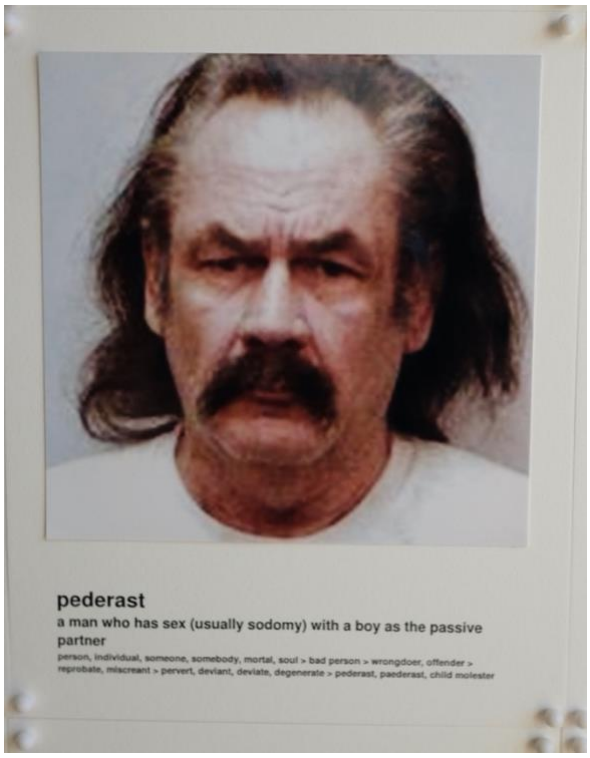


Figure 8 & 9: *Two children labeled as bastards and a man labeled as pederast. From the dataset ImageNet, as presented in* Training Humans. *Photo: Tuva Mossin.*

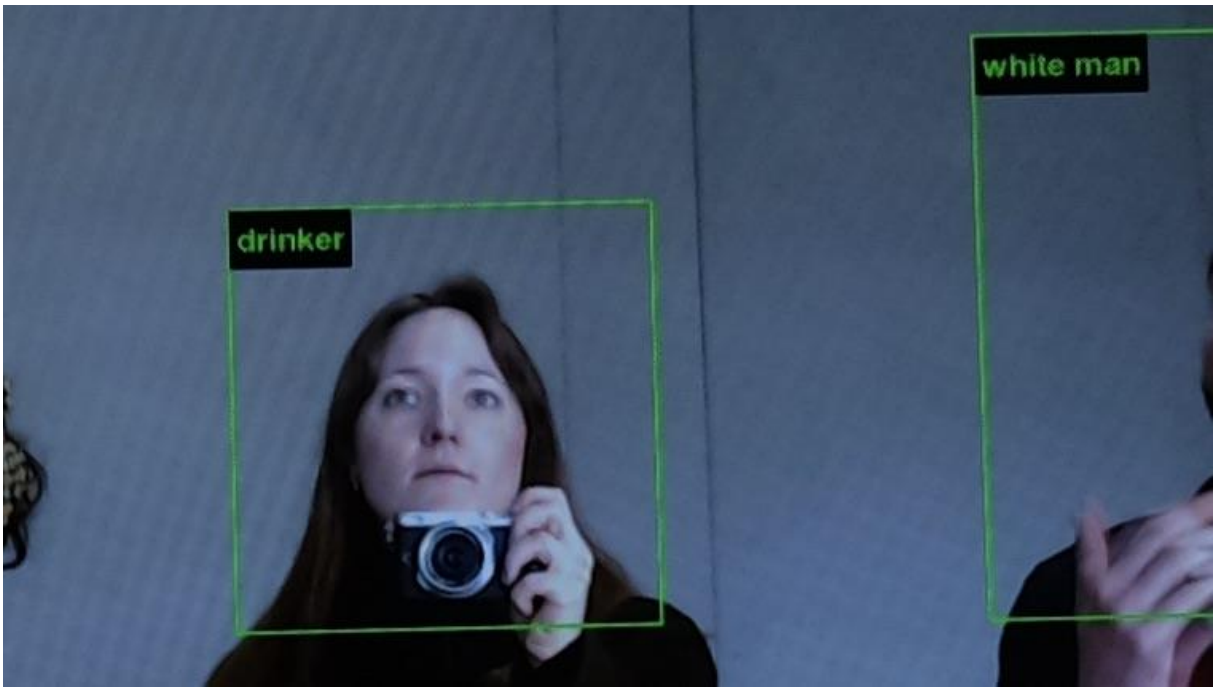


Figure 10: ImageNet Roulette *as interactive installation in* Training Humans. *Curiously, whenever I took up the camera in order to photograph the screen, the tag on my face changed from "psycholinguist" to "drinker". Photo: Tuva Mossin.*

Figure 11: ImageNet Roulette *as interactive installation in* Training Humans. *Photo: Tuva Mossin.*

Documents of a Visual Practice

Up to this point, I have described how Trevor Paglen and Kate Crawford set out to scrutinize the training image dataset ImageNet and how this project resulted in an online app, an online essay and an exhibition – all of which were published in September 2019. In addition to these, the duo also published a short catalogue accompanying *Training Humans*, as well as two smaller exhibitions which have not been discussed.[10] In different ways, all of these works serve to raise concerns regarding automated ways of analyzing people. Amongst these, the most obvious concern the biases embedded in ImageNet at the level of its taxonomic basis – along with the practices of collecting and managing these digital archives. On a more abstract level, we find questions concerning the contexts in which these technologies are embedded, as well as matters of representation – in both senses of the word. At this point of the analysis, however, I am interested in taking a step back to discuss the project as a whole: what techniques does Trevor Paglen employ to visualize practices which are, even though concerned with vision, not in themselves visible in the optical sense?

[10] I am here referring to the exhibition *From "Apple" to "Anomaly"* shown at the Barbican Center in London (Sep. 2019 - Feb. 2020) and *Making Faces* in collaboration with Prada Mode at the Maxim's restaurant in Paris (Jan. 19.-20. 2020). Of these exhibitions, however, *Training Humans* appears not only to be the most extensive presentation of the subject at hand, but also the one which the duo themselves, according to my own observations, has promoted most avidly.

While engagement with *ImageNet Roulette* was essentially a personal experience – probably enacted by most on their computer or smart phone in the private comfort of their home – the wider context of its reception expanded to the public sphere. In this way, millions of people not only got an intimate experience of how their selfies became implicated in a system categorizing them based on physiognomic traits. Additionally, the myriad of social connections created through #ImageNetRoulette also facilitated a kind of spectatorship where the individual became part of the work itself whilst simultaneously enabling a view of the app as visual *practice* on a systematic level. By this I mean that the biases and discriminatory practices are first obvious when applied to a multitude of differently labeled images which are shown in relation to each other – as they were on social media and other news channels. Only then did the roulette go from producing mere glimpses of machine vision into becoming a document of a discriminatory practice per se. A counterargument to this could be that the software embedded in the app was only *trained* in ImageNet and not an inherent part of the database itself, and that *ImageNet Roulette* therefore was not an actual representation of the database. But, as the essay *Excavating AI* carefully argues, these biases do point back to the taxonomy of the database itself. The exhibition served to hammer home exactly this point.

Furthermore, the project facilitated a sense of conflation of the private and public spheres – a quality typically attributed to activity on social media (Trottier 2016: 68). While this tendency is hardly surprising, *ImageNet Roulette* was, for many, probably the first and only tangible experience thereof. Even so, I want to point out that *ImageNet Roulette* was arguably a more than averagely abstract representation of its subject, meaning that it was not at all intuitive and needed quite some explanation. Although this is not a problem in itself, it is thinkable that this may obscure the connection between the artwork and the training database it represents to the general audience. Perhaps this is why Paglen and Crawford chose a different strategy of representation in the exhibition *Training Humans*. Here, the various databases are all represented according to the same general formula: a selection of photographic portraits – either still or moving – hand-picked in order to represent the body of images that make up the dataset as a whole. What I find interesting about this is how the presentation of the photos according to grid-like structures functioned to make the digital archives appear as documents of a *practice.*

Lastly, one could ask what the exhibition added to the discussion which had not already been made explicit through the app and essay. My first impression when there, was that in being so heavily reliant on contextualization, it at times felt more like visiting a natural history museum than an art gallery. Also, as the above analysis makes clear, the viral app gained so

much attention that when I finally got to Milan, I was already well acquainted with the project. Thus, the exhibition was left somewhat overshadowed by the app, which made me question whether traditional, "analogue" exhibitions are simply too slow in format compared to online phenomena.[11] In this sense, the trip to Milan was somewhat anticlimactic in that it did not live up to the rush and feelings of *extreme present* facilitated by *ImageNet Roulette* – a quality which "star curator" Hans Ulrich Obrist predicted in 2016 would be characteristic of art of the future (Obrist 2016). Admittedly, the exhibition did add an atmosphere of intense monitoring which was not (for my part at least) a quality that could be ascribed to the app. In this sense, the curatorial strategies in *Training Humans* succeeded in establishing an affective bond between me and the faces staring out at me from the walls, as if we were all part of the same group. Thus, besides presenting an overview of an evolution of TIDs, *Training Humans* also appeared to represent a certain unity of people across the globe – of all those subjected to surveillance infrastructures. While looking at the various works in the Osservatorio, I had a strong sense that even though my own picture was not present, it might as well have been – and was indeed in the final room.

In addition to the above observations, I would add that the multitude of representational techniques employed in ImageNet Roulette effectively functioned to blur the boundaries between physical and digital space. Having experienced instances of this multifaceted art project, it is apparent that the latter is indeed an integral part of the former, and that the commonplace way of referring to "normal" space as *physical* – as if this would exclude the digital – is misinformed. This, in turn, gives rise to several questions regarding the relationship between gallery spaces, whose basic structures have not changed notably in the last few decades, and the presentation of digital art in them. What I find interesting about *Training Humans* in this respect is how, when ascending to the second gallery floor, a chronological leap to the age of Internet 2.0 was immediately apparent. There are two factors, I believe, that played an especially important part in forming this impression. The first is the dramatic increase in the number of images presented, and the second factor is a shift in their aesthetic quality. By this I mean that while the images presented on the first floor were highly stylized and clearly produced for research purposes, the images along the walls on the second floor were so-called snapshots. In light of this, I ask whether it is meaningful to understand the subject matter in

[11] In light of the ongoing pandemic during which the art world has seen an ever-increasing number of born-digital exhibitions, this question seems all the more relevant than just months before.

Training Humans not simply as a history of training sets, but also as an exhibitionary projection of what is now termed an *ecology of images*.

A New Media Ecology

In the book *Digital Snaps*, Jonas Larsen and Mette Sandbye tell of how, at a conference in 2009, visual culture scholar Martin Lister called for the establishment of what he termed a *new media ecology* in order to make sense of a rapidly changing media reality brought about by digital photography.[12] As a response to this, the editors published a collection of studies by various scholars within the visual studies meant to reflect upon exactly this newly evolved media reality. The book's aim, the editors stated, was to establish a novel paradigm concerning digital snapshot photography in particular, highlighting new technologies, meanings, circulations and practices. Referencing Lister, Larsen and Sandbye also referred to this new paradigm as a media ecology,[13] the central thesis being that technologies can and should not be separated from their contexts and embodied everyday practices: "photography must be understood simultaneously as a social practice, a networked technology, a material object and an image" (2014: xxiii).

In his contribution to *Digital Snaps*, Lister emphasizes how digital snapshot photography – which today makes up the vast majority of photographs – has become a constituent part of rapidly evolving new ways of seeing, doing and valuing photography (2014). In light of the increased accessibility of cameras facilitated by smartphones, he argues that photographs are now taken more lightly than previously and, in all probability, deleted just as readily. Many digital snaps, Lister suspects, remain as code in hard drives or "clouds", continually in flux, reducing their visibility to a *potential* rather than a fixed state.[14] If actually viewed at all, he notes, it is often just as "temporary emissions of light" on screens that are also used for a multitude of other purposes (Lister 2014: 1). The consequences of these developments are manifold. One is that the digital snap has become reciprocally intertwined with the internet, and consequently ingrained in a web of user generated content, serving commercial interests. Importantly, these changes to the status of photography are typically

[12] "Maybe what's happening now for photography was always its destiny and fate. But it's not the end of photography. It's rather the end of photography as we know it. To understand this change, we need a new media ecology." Cited in Larsen and Sandbye, *Digital Snaps*, p. xv.

[13] It should be noted that use of the ecology metaphor within media theory can be traced all the way back to the work of Marshall McLuhan in the late 1960s. What is supposedly new here is the specific attention towards digital photography within such a framework.

[14] As opposed to, for example, developed analogue photographs stored away in a shoebox, which are inherently visible even if no one is looking at them.

presented in what Lister describes as "sweeping, large-scale terms; they are met on a scale of almost unimaginable quantity, a kind of digital sublime" (2014: 10).

As is typical when facing things of sublime dimensions, the subject at hand is to grasp, resulting in a state of confusion or worse. This is a concern expressed by Lister:

> We appear to live in an information age in which there is, paradoxically, a historically unprecedented loss of information while we anxiously weigh it in gigabytes and worry as it quite literally heats up our computers and has us digging ever bigger cables into the ground (so much for the immateriality of information). It is in this context that I think there are grounds to argue that we are drowning in images, documents and artefacts, that we have too much stuff. The sheer mind-boggling quantity means that the game is up (2014: 15).

The fear that the quantity of images that exist today can lead to the kind of overwhelming helplessness expressed by Lister was echoed by the art historian Tom Holert three years later. He points out how this exact tendency can be seen in contemporary art more generally – especially in relation to art concerned with surveillance – arguing that engaging with big data and notions of the mathematical sublime may effectuate certain "aesthetic/existential stances such as horror" (2016: 277). It is a proximity to this kind of mathematical sublime, in addition to the snapshot qualities, I believe, that render the second-floor images in *Training Humans* so recognizable as photographs scraped from the web. Importantly, the experience, for my part at least, did not tip over to become fully sublime. This is good because the sublime is usually associated with a loss of words, prohibiting clear thinking in relation to whatever phenomenon induced it. However, the near-sublime experience should not be completely reduced to a matter of quantity, as it was also a question of tempo. I have already described how the pace at which the TV-screens in the exhibition quickly skipped from image to image gave the impression of witnessing an automated alternative to looking or observing: while I was not even able to have a good look at every photograph in the exhibition, facing these screens made it clear how a computer system would probably be able to scan them all within minutes, if not instantly. This made me ask whether we have come to a point where we are so dependent on AI that we can no longer keep up. This is how I understand Lister when he argues that "the game is up."

Furthermore, it goes from the above analysis that when presented collectively as representations of training sets, the digital snaps in *Training Humans* did not only give rise to uncanny feelings, but simultaneously spurred a whole set of serious questions. Just like the wall text promised, it was obvious from looking at the photographs in *ImageNet* and *Training Humans* that this kind of user generated content is used to develop technology to *look back* at us. This is in line with Lister's argument, as he claims that interactivity is key for understanding

the new modality of snapshots which he is trying to pinpoint: as creative users, we are always contributing "at the expense of the surveillance of our habits, choices, predispositions and patterns of consumption" (2014: 8). In this respect, he suggests that thinking of these changes in terms adopted from the natural sciences can be useful:

> Ecology is a useful word here, with its roots in a concern for understanding the pattern of relations between organisms and their environment. It draws out attention not only to changes in an environment, but also the adaptations and mutations undergone by an organism or species as it lives on in changed circumstances. It captures something of the complex interplay between emerging media and communication systems and behaviours, the convergence and recombination of old and new media forms, and the remediation of the old by the new (2014: 16-17).

In the context of *Training Humans*, I find this way of thinking useful because it helps to put words to my experience of ascending not only to a new space on the second gallery floor, but simultaneously of entering a new media era. However, in contrast to how Lister first described it roughly a decade ago, I did not experience this media ecology as new *at all* when encountering representations of it in the 2020 exhibition. On the contrary, the idea of photographs as an ingrained part of internet culture was experienced by me as a commonplace fact – hence the immediacy of the association.

My perspective on this might, of course, be informed by my age and personal experience: nine years younger than the World Wide Web (1985) and four years younger than the first commercial version of Photoshop (1990), I have almost no memory of ever not having owned a digital camera or my own personal computer with internet connection. Uploading and sharing photos on sites such as Tumblr and Piczo has been a normal way for me to socialize and perform identity online since my early primary school years, before getting a Facebook account in the 8th grade (2008). Thus, photography as something essentially digital, social and potentially networked, is an idea just as readily available to me as the idea of the medium as analogue. Any feelings of newness when facing the second gallery floor were thus only relative to the older material on the first floor. In light of this, I find the most perplexing aspect of ImageNet Roulette to be just how long it took for a project like this to arrive. I wonder if it is this same concern Damn Magazine journalist Sophie Rzepecky has in mind when she comments about *Training Humans* that "the questions provoked by this exhibition are urgent yet lag behind the extremely fast rate technology is already being live tested on populations" (Rzepecky 2019). As these problems are clearly no longer new in any meaningful sense of the word, one might wonder if traditional art- and academic institutions work at a pace too slow to keep up with digital technologies evolving at an exponential rate. This is one of the many dilemmas facing anyone

seeking to critique developments within the tech business, as has been noted by Tom Holert in the 2016 exhibition catalogue *Watched! Surveillance, Art & Photography*: "Seeking to keep track of the development of social algorithms, biometrics, sensor technology or satellite reconnaissance often proves to be a futile endeavour, with belatedness being a potential default position in many a practitioner's process"(2016: 276).

In the same catalogue, art historian Louise Wolthers resumes the discussion from *Digital Snaps*, arguing for the importance of developing critical perspectives on what she calls *viewing ecologies*: as situated within a complex rhizomatic[15] network of visibility, she stresses the need for critical tools by which to navigate and make sense of the contemporary media reality we find ourselves in (2016: 283). This, I think, is of special importance precisely in order to avoid the kind of helplessness feared by Lister and Holert. Furthermore, Wolthers emphasizes how the responsibility weighs extra heavily on theorists and artists in privileged positions, in that not everyone is affected equally by surveillance practices. Together, we need to develop a kind of "environmentalism of the image," she argues, "understood as a critical awareness of the over-consumption of images, an excess of photography, and mutual mass surveillance" (Wolthers 2016: 284). While academics coming from social studies typically work to look back and expose surveillance practices, it is imperative that artists and visual theorists strive to find methods that "counteract such practices by identifying image and gaze ecologies that draw on the possibilities of analogue photography (…) as well as on characteristics of the digital image" (Wolthers 2016: 285). This necessitates taking a step back to find spaces where one is able to see oneself as part of an ecology; a system we are not only part of, but also mediates how we see and relate to the world around us (Wolthers 2016: 285).

Drawing on Ariella Azoulay's theory of photography as a civil contract, Wolthers clarifies that this call is not only about the right to see, but also "the right to enact photography free of governmental power and even against it, if it inflicts injury on others who are governed" (2016: 288). If applying this theory to ImageNet Roulette, I would conclude that this project serves a double role by taking on the responsibility of the social scientist in exposing a highly problematic surveillance practice, whilst simultaneously creating the kind of space where we are able to perceive ourselves as part of such a system. This is true of the app and the exhibition alike.

---

[15] I have noted that Wolthers, as well as many others writing on the topic often use terms like "rhizomatic" in descriptions of media landscapes or ecologies. I take this to refer to Gilles Deleuze and Felix Guattari's philosophical concept based on the botanical rhizome, first developed in *Capitalism and Schizophrenia* (1972-1980). This particular conception of a media ecology can thus be understood as situated within a wider discussion of knowledge production.

A Nascent Paradigm Shift?

In one video promoting *Training Humans*, Paglen argues in favor of approaching AI in machine vision through a practice of looking at images, or even as a form of photography criticism (The Photographers' Gallery 2019b). If taking this claim at face value, one might understand the exhibition not only as a criticism of the databases involved in AFR, but also as a scrutiny of a visual practice on a more general level. In a broad sense, the project can thus be understood as a space where two practices, art and machine vision, are contrasted against each other. As I understand it, ImageNet Roulette engenders the exact kind of critical viewing ecology that Wolthers called for. In this respect, I find it especially interesting that this is not achieved by taking "a step back", as she put it, but rather by diving headfirst into the materiality of the dataset itself. Maybe this specific approach is a result of Paglen and Crawford's very different backgrounds? Coming from different disciplines, art and geography (Paglen) and critical technology studies (Crawford), the duo found themselves asking many of the same questions even before they started working together. Moving on, the two discuss how Paglen worked with images in order to create *new ones*, while Crawford was working with them in order to understand the epistemologies, politics and philosophies that underlie the construction of datasets. When they started collaborating, they saw how these questions all started to converge, how the materiality is actually part of the philosophy and epistemologies, and vice versa (The Photographers' Gallery 2019b). This takes us to the last topic I want to discuss in relation to Image Net Roulette before rounding off this chapter: how does the project reflect the role of the *artist as a producer of knowledge*?

When discussing this question at an earlier phase of my MA research, I was critical of what I then considered an overdependence on text to back up the projects – especially in the case of *Training Humans*.[16] True, as with every other art exhibition, a considerable amount of background knowledge is essential for grasping the full depth of its meanings. This is probably why Paglen and Crawford have taken such care of producing various texts and videos in order to contextualize their own work. Yet, one cannot expect the general audience to read the full length of these long and, at times, rather challenging texts – and even if one actually takes the time to do so, there are still many questions left unanswered.[17] For example, the essential

[16] I refer here to an unpublished seminar text written by me in qualification of the course KUN221 held at the University of Bergen, spring 2020.

[17] As a basis of understanding the amount of work this involves, I (who is acquainted with the relevant vocabulary) spent around one hour reading the text Excavating AI, and around thirty minutes reading the catalogue text (which is largely a reformulated repetition of the same content). If including all the tweets and news articles informing the general reception context, even more hours must be estimated.

question of what contexts or situations AFR technologies are actually embedded in, is not clearly answered in any version of the project. If understood as a form of artistic activism striving to spread awareness of the possible dangers of allowing unregulated use of such technologies, the project as a whole thus has some notable setbacks. Upon further consideration, however, I am wondering whether this premise is misinformed. As outlined before, it might be more fruitful to think of ImageNet Roulette not only as a form of activism, but as a series of critical comments meant to contribute to exactly the kind of critical environmentalism of the image called for by Wolthers. This conception is supported by the fact that Crawford on her personal webpage refers to ImageNet Roulette as a "multi-year research project."

Having had much time to watch films during the ongoing pandemic, I stumbled upon a documentary called *The Edge of All We Know* (2020). It showed how a network of more than 300 scientists managed to collect and systematize an overwhelming amount of telescopic data in order to prove the existence of a supermassive black hole in the galaxy Messier 87. This feat was accomplished with the use of the Event Horizon telescope; an assembly of separate telescopes across the globe, together simulating one massive Earth-size telescope. Through this film, I came to learn that because black holes are not actually observable in themselves, the scientists instead relied on extremely complex visualizations of the area *surrounding* it. Thus, it was the images of light from the stars in orbit *around* the black hole (the event horizon), and not the hole itself that allowed these scientists to prove its existence in April 2019 (Fig. 12). The reason I bring this up, is because it is my understanding that just like the black hole, the visual practices revealed by ImageNet Roulette – even if concerned with optics – are also not inherently visual phenomena. Analogous to the vast amount of differently trained scientists who collaborated in order to reveal the black hole, Paglen and Crawford's project can be understood as situated within an interdisciplinary field in dire need of techniques for visualizing the phenomena it intends to scrutinize. The method the two refer to as an *archeology of datasets*, can in this sense be understood as analogous to the ecology proposed by Lister and Wolthers, although with somewhat different connotations.[18]

[18] Etymologically, the word points towards ancient history, which is fitting as the text is concerned with a kind if evolution. The word's similarity to the word *archive*, while not etymologically related, may also point towards a logical connection between the two. Of course, the phrasing can also be interpreted as a reference to Michel Foucault's methodology. See e.g., Foucault, *The Archeology of Knowledge* (1969).

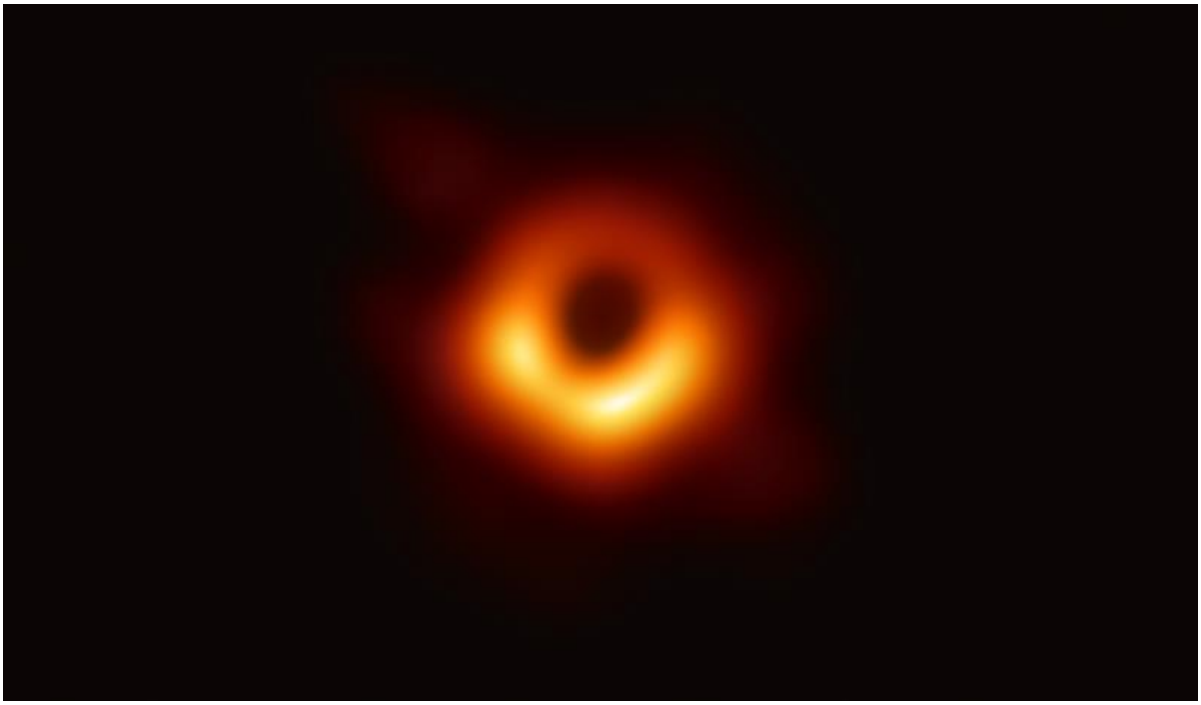


Figure 12: *This image was presented by the Event Horizon Telescope Collaboration at six simultaneous press conferences worldwide 10.04.2019 proving the existence of a supermassive black hole at the center of M87. Retrieved on 28.01.2021. URL:* *https://eventhorizontelescope.org/press-release-april-10-2019-astronomers-capture-first-image-black-hole*

As is increasingly being argued by practitioners within the field of machine learning itself, its methods are becoming more and more reminiscent of alchemy. Without the tools offered by scientific method, alchemists relied on recipes and formulas of how to achieve certain reactions without possessing actual understanding of the chemical processes involved. It is along these lines of thinking that certain critics within the field have recently started to argue that even when the *results* achieved with AI are satisfactory, it still constitutes a problem that the processes involved are essentially opaque (*the black box problem*). One example of such a critic is the Google AI researcher Ali Rahimi who, at a conference in 2017, argued that machine learning is lacking scientific rigor(Rahimi 2017). According to him, researchers are implementing and sharing techniques that they ultimately do not understand – just like the alchemists did until the advent of scientific theory forced them to throw away thousands of years' worth of accumulated know-how. Admittedly, he adds, there was a place for alchemy in its time: "alchemists invented metallurgy, ways of dyeing textiles, our modern glass-making processes, and medications. Then again, alchemists also believed they could cure diseases with leeches and transmute base metals into gold" (Rahimi 2017). At this point in history, where AI has become deeply ingrained in essential parts of society, the field is too important to be treated this carelessly. As Rahimi emphasizes, this attitude might perhaps have been ok when building

simple photo sharing systems, but "we're beyond that now. We are building systems that govern healthcare and mediate our civic dialogue. We influence elections" (2017).

As mentioned in the discussion of *Excavating AI*, there is a tradition within the AI community to handle problems almost exclusively in technical terms. While this has been criticized by people such as Hubert L. Dreyfus at least since the early 1970s (as will be discussed in Chapter 2), it has – at least to my knowledge – been mostly disregarded by those with the power to change the course from within the field itself. There is thus an irony at play here in that it took for this kind of technology to become ingrained within so to say all parts of societies across the globe before issues as those discussed in this chapter are starting to be taken seriously. When this is finally being done, it is apparent that the tech developers themselves do not seem to have in place neither the vocabulary nor the methodology needed to handle it properly.

Although Rahimi's critique has been contested by some who do not see the above problems as serious, it is obvious – at least to someone coming from the field of art history – that we are here faced with a challenge that requires broad interdisciplinary collaboration. As Wolthers puts it in *Viewing Ecologies*, "the spread of photography, the act of photographing, and the circulation of photographs demands the same degree of responsibility as freedom of speech" (2016: 286). At a time when both the EU and the US Senate are taking action to limit the extreme level of control exercised by Silicon Valley's tech-giants through communication platforms, it is high time that the same amount of attention is being directed towards the visual aspects of this. It is my interpretation that projects such as ImageNet Roulette is just what might give us the glimpses into the black boxes of AI that are needed in order to bring about such changes. That the project was a collaboration between an AI researcher – who co-founded a university institute specifically in order to research the social implications of AI – and an artist fellow who is also a geographer, is only typical of this flourishing sector of cultural production.

Conclusion

One of the questions asked in *Training Humans* was if "the boundaries between science, history, politics, prejudice and ideology in artificial intelligence" could be located. After having analyzed different versions of the project, it is clear that no such boundaries exist. From the aesthetic perspective offered by ImageNet Roulette, the quest to "map out the entire world of objects" seems just as mystical as setting out to find the Philosopher's stone. At the same time, the project makes clear to the audience how, besides a list of epistemic issues, the machine vision industry has several other dark sides, such as mass-harvesting of images online and

exploitation of underpaid online workers through platforms such as Amazon Mechanical Turk. If we try to understand ImageNet Roulette as an environmentalism of the image, it is apparent that the kind of interdisciplinarity exemplified by Trevor Paglen and Kate Crawford is essential. Furthermore, the analysis in this chapter is illustrative of the fact that these problems are not only technical, but material and epistemological as well. In light of this, it is not the art that is overly dependent on academic backup to prove its relevance, but rather the field of AI that needs intervention from outside in order to make sense of the industry's own blind spots.

ImageNet Roulette is undoubtedly a big step in the right direction, but in light of technological developments, one can nevertheless question its efficacy as it appears to lag years behind. In saying this, I echo Holert and Rzepecky in worrying that the publication of ImageNet Roulette is simply too little and too late. What is perhaps the most frightening aspect of it all, however, is the fact that ImageNet and the other TIDs presented here (having been built for research purposes at a university) are only some amongst a few which are open to the public. Most AFR systems are hidden in ways that render them resistant to scrutiny. There is thus a chilling paradox at the core of this analysis in that the TIDs discussed here are probably not even the most problematic ones. This being said, it would be unfair to expect of a single art project to give answers to complicated questions which are rather belatedly starting to be asked. I will return to this question in Chapter 4 after having discussed Heather Dewey-Hagborg's project *How do you see me?* in Chapter 2, and Broomberg & Chanarin's *Spirit is a Bone* in Chapter 3. Analyzing all three projects together from an (art)historical perspective, I hope to counter this attitude by arguing that aesthetic approaches towards issues of surveillance might be especially promising given the special kinship with art in terms of visibility and/or opacity.

Chapter 2

The Treachery of Facial Recognition: *How Do You See Me?* by Heather Dewey-Hagborg

Introduction

Building on the previous chapter, this discussion will center around an art project by the American artist Heather Dewey-Hagborg (b. 1982), who is also an artist fellow at the AI Now Institute at NYU. Describing herself as an artist and "biohacker", Dewey-Hagborg fits the description of the transdisciplinary artist typical of surveillance art. Holding a master's degree in interactive telecommunications from NYU and a PHD in electronic arts from Rensselaer Polytechnic Institute, she is interested in exploring art's potential as research and technological critique (Dewey-Hagborg n.d.). Perhaps most widely known for the project *Stranger Visions* (2012-13), which included sculptural portraits made on the basis of found genetic material, Dewey-Hagborg has for several years created highly interesting projects in the intersection between art, science and technology. In Norway, her work *Probably Chelsea* (2017) has been exhibited both in Bergen at Kunsthall 3.14 in 2018 and in Oslo at Kunstnernes hus in 2021. This was a series of portrait sculptures made on the basis of genetic material sent to her from the then imprisoned whistleblower Chelsea Manning. On a general level, all of Dewey-Hagborg's recent projects can be said to revolve around technologies used for surveillance. The project *How do you see me?* nevertheless stands out in that she has moved from focusing on the biological foundations of forensic techniques and other cultural phenomena to an analytical engagement with an AFR algorithm.

How do you see me? was published in September 2019, commissioned by The Photographer's Gallery in London as part of a year-long digital program exploring the "ever-evolving role and effects of scientific image datasets" (The Photographers' Gallery 2019a). As a response to the present saturation of AFR technologies in both private and public life, Dewey-Hagborg was concerned about what we really know about these systems. For her, *How do you see me?* constituted a way of "looking back and trying to learn how this alien intelligence that is so attentive to my every move is structured, internally" (Dewey-Hagborg 2019). In order to achieve this goal, the artist used what is called *adversarial processes* to deceive machine learning systems to reveal the limits of what can be recognized as her face. These are filters that can be applied to digital images, leaving them unchanged to the human eye while posing serious difficulties for object detection algorithms (Zhou & Firestone 2019). In addition, the artist used

computational tools inspired by evolution to generate a population of images where the ones deemed most easily recognizable as her face got to propagate (The Photographers' Gallery 2019a).

Inspired by the Darwinian metaphor of survival of the fittest, the images, over hundreds of generations, evolved into abstract forms that the system detected as Dewey-Hagborg's face even though they to the human eye look nothing like her – or any face at all. In this sense, the project can be understood as a kind of experiment seeking to uncover what the algorithm is actually looking for in faces. From there, Dewey-Hagborg's process mainly consisted in curating the images, picking out those that served her intellectual and aesthetic goals. The result was two short videos as well as a series of stills, of which half focus on the *detection* of faces in general, while the other half focuses on the *recognition* of her face in particular. These were presented as an installation on the so-called Media Wall at the Photographer's gallery in September 2019 (Fig. 13), but as I was not able to visit this exhibition, the following presentation is limited to the excerpts available on the artist's homepage as well as other online documentation.[19]

The overarching question guiding the discussion in this chapter will be the same as the work's title: how do facial recognition algorithms "see" the artist, and by extension people in general? Answering this requires a general presentation of the technical processes involved in the recognition and labeling of faces from original input to final output. From this discussion, it becomes clear that the notion of *intelligence* implied by the term artificial intelligence is misleading. This leads me to take a step back to look at the history of AI before discussing whether it makes sense to separate understanding from bodily experience. Here, I will draw on the work of philosopher Hubert L. Dreyfus, who already in the early 1970's criticized the notion of intelligence in machines in his work *What Computers Can't Do*. As AFR is widely considered to be at the very frontier of AI, this discussion seems pertinent.

Presenting The Work

The screen is all black, but within seconds, a field of white color gradually spreads across the background in a cone-like formation, continuously shapeshifting, morphing. The shape appears almost like a beam of light emitted from a flashlight, searching. Suddenly, there is a shift: a small, almost triangular white area is singled out in one corner of the scene and a green square

[19] You can read about the project and see a video excerpt on Dewey-Hagborg's personal webpage. URL: https://deweyhagborg.com/projects/how-do-you-see-me

appears in the middle of it. The image fades out but leaves the viewer with little time to reflect on what she has just seen because in a split second, it is filled again. This time, the viewer is presented with what appears to be a collection of stills from the same kind of clip: grids of separate squares depicting abstract black and white forms are now presented in short, subsequent flashes. Again, there is no time to single out any one image to focus on before a new assemblage is presented. After a while, one frame depicting a white oval is framed by a green square. This still then remains frozen in the upper left corner while the rest of the images continue to be in flux. Then another still is frozen, framed in green. Then another, and another. By the time the series of assemblages fade out, five stills have been singled out. The screen turns black before we are once again presented with the same kind of moving image flow as starting out: a white shape morphing across a black background. Then, in quick succession, three green squares appear. The image fades, and the clip is over.

Based on the reviews I could find, this film – although only an excerpt – appears to capture the general gist of how it would have been to see *How do you see me?* as it was exhibited in London. The reason I emphasize the sense of *seeing* is that there is no sound accompanying the video. This gives the experience of watching an alienating feeling and emphasizes how the purely optical nature of machine vision differs from a human, situated experience involving the whole body. Furthermore, the film experience supports the perception of facial recognition systems as what Dewey-Hagborg describes as “alien intelligence”. The stills by themselves, however, do not appear as alien. While the film gives the impression of watching some kind of process, the stills don't give away any clues to what they are or how they came about (Fig. 14). Had it not been for the green squares and blue circles, they could have passed either as overexposed black and white photos, or bright flashes of light in the night. Some of the stills depicting curved shapes even resemble how I would imagine digitized and saturated minimalist paintings. So how are we to understand these images?

As Dewey-Hagborg explains it, she broke the problem of making this project down into two parts: 1) the detection of faces in general, and 2) the recognition of her face in particular (Dewey-Hagborg n.d.). This makes sense, as these are the two most basic functions in any facial recognition system. The first group is the images indicated by blue circles, while the second group is indicated by green squares. What I saw in the video, then, was the system "seeing" the artist's face amongst the many frames of abstract shapes, giving the viewer a glimpse into what the facial recognition system deems to represent a face. Reflecting on this, Dewey-Hagborg explains that during the process, a certain pattern became apparent:

> What comes up over and over and over again in many generations, in many runs of the algorithm, is this kind of white oval. That is very revealing about these kinds of stereotypes that are embedded within facial recognition and facial detection technology. This is a very direct and very political kind of connection to the context of the dataset (The Photographers' Gallery 2019a).

In light of the discussion in the previous chapter, it should be clear that the stereotypes referred to by the artist here point to a profound problem in terms of representation in both senses of the word. But what is even more interesting in the particular case of *How do you see me?,* is how the images demonstrate that the algorithm clearly has no understanding of what a face actually is. This goes for the detection-images as well, although the shapes here are more complex, and as the artist herself describes it, more "poetic." As Dewey-Hagborg puts it, it is visible that "there's these totally strange vectors within the system that are neighboring to the vector that represents my face that makes no sense to us" (The Photographers' Gallery 2019a). Thus, even though there is seemingly some kind of logic to the recognition process, the project still reveals that this logic is something entirely different from an understanding of what a face is or means. Furthermore, the artist asks, "what would it even mean that there are these latent possibilities lurking in there that something completely different than my face can also be just as well recognized as me?" (The Photographers' Gallery 2019a).

While *How do you see me?,* like ImageNet Roulette, certainly raises questions of both authority and power, this is not the discussion I will focus on in this chapter. What from my perspective makes this particular project stand out from other artworks concerned with AFR is the special door it opens for a discussion of the concept of understanding in machine learning, and thus also notions of intelligence. This reminds me of a passage in *On Photography* where Susan Sontag questions the assumption of photography as a record of truth: "photography implies that we know about the world if we accept it as the camera records it. But this is the opposite of understanding, which starts from *not* accepting the world as it looks. All possibility of understanding is rooted in the ability to *say no*" (2010 [1977]: 23 [italics added]). As facial recognition by machines is essentially a process of matching an identity to a face via images, this point is of crucial importance when trying to understand such technologies. While, as Sontag also admits, photographs can be *instructive* of something of which one does not already know, this is not the same as understanding. Considering the fact that object recognition systems' only connection to the external world is *through photography*, this point is of crucial importance. But what does this non-allegorical platonic cave situation mean? And what are the consequences?

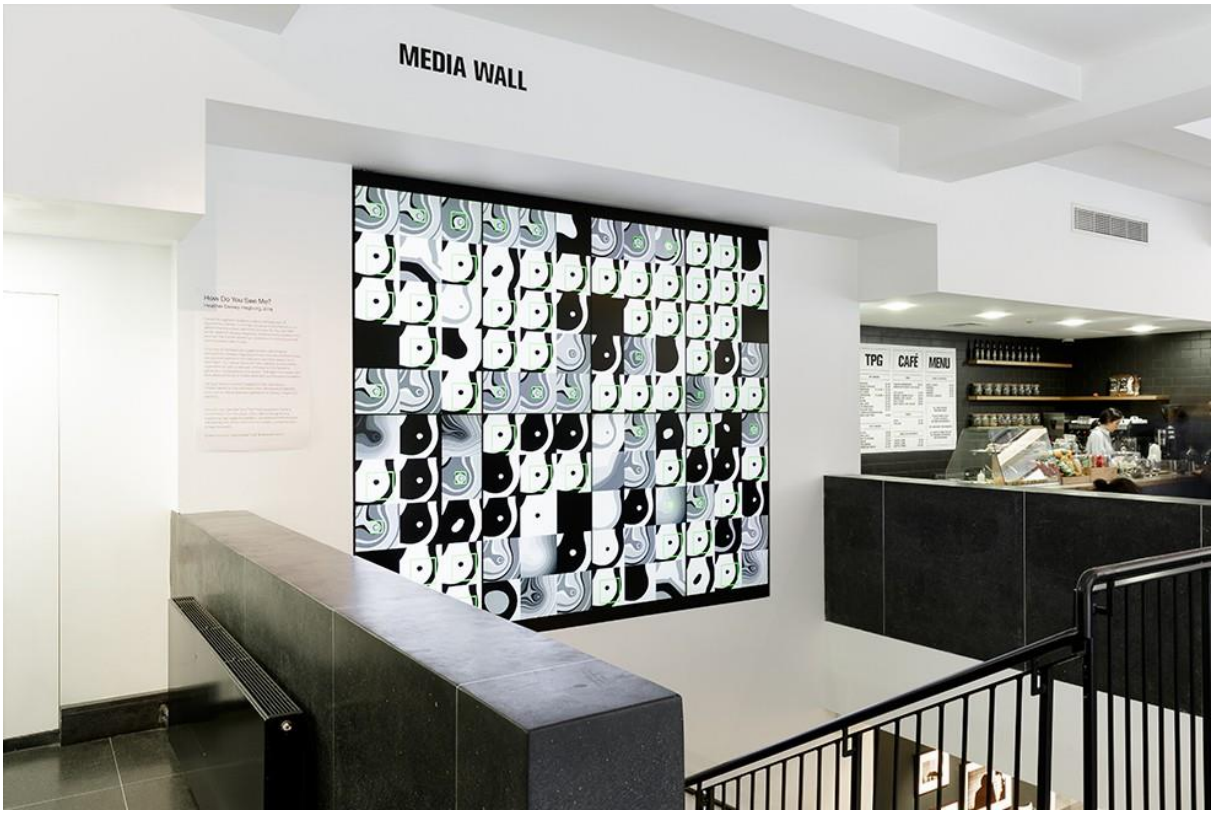


Figure 33: How do you see me*? As presented at the Photographer's Gallery. Photo: Tim Bowditch. Retrieved on 19.02.2021. URL: https://unthinking.photography/articles/im-looking-at-you-looking-at-me.*

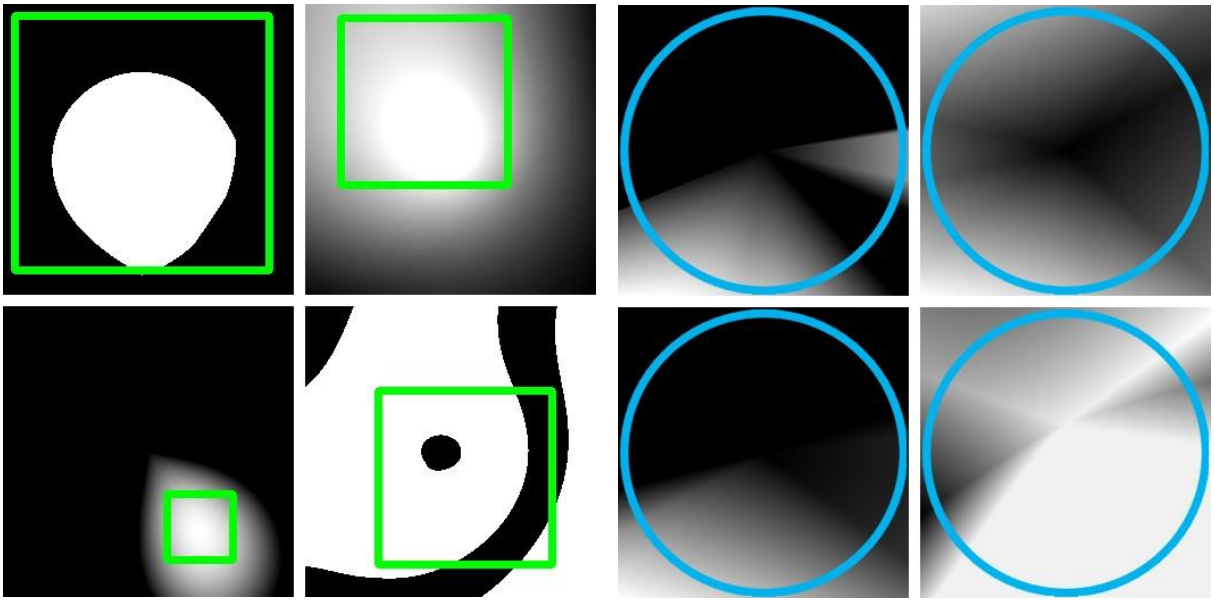


Figure 14: *Stills from Heather Dewey-Hagborg,* How do you see me? *(2019). Retrieved on 19.02.2021. URL: https://unthinking.photography/articles/im-looking-at-you-looking-at-me*

Layers of Abstraction

I have already stated that facial recognition systems can be understood as the matching of faces to identities via images. On a technical level, AFR works in just the same way as any kind of automated object recognition. Historically, this task has been handled in many different ways, but today, *convolutional neural networks* (CNNs) are the preferred option. These computational models are inspired by how the human brain processes information and have been proven especially efficient for object recognition. In the following, I will describe the basic structure of neural networks in general before moving on to CNNs specifically. I will not go into detail on the specific mathematical functions involved, as I deem it sufficient for the discussion at hand to be acquainted with the general architecture of neural networks.

The most basic component of any neural network is the node[20], of which there can be one or more layers. On a general level, each node contains a function *(f)* which computes a *weighted* output based on the values of inputs from the previous layer. For example, a given node *f1* can get two inputs X1 and X2, which it uses to calculate its output Y. This also includes the weight *b* (bias), to which I will return shortly. The most basic neural networks are the *feedforward* ones, in which the direction from input to output only goes one way. While these *can* have only one layer (input and output), a slightly more complex feedforward network is more common because it can learn non-linear functions as well as linear ones (Karn 2016b). This is usually preferred in image recognition, as it supposedly captures the complexity of input data better.

Such a *multilayered perceptron* (MLP) contains one or more *hidden* layers that have no direct connection to the outside world. To explain this, I will present an example of an MLP containing three layers (Fig. 15): the first layer contains three input nodes, the second (hidden) layer also contains three, while the third layer only has two nodes. The first layer is the one that connects to the "outside world". In our example, that would be data contained in an image, such as pixel values. The nodes in this layer do not perform a calculation, but simply passes on the information (e.g., the color of each pixel) in the form of a number. In the next layer, then, each node performs a separate calculation based on the data it has received, before also passing it along. The *output nodes*, which are contained in the outermost layer are responsible for final summations, which are communicated to the outside world in one form or another (e.g., a green square, a label or a number).

[20] These components are also commonly referred to as *neurons* or *units*.

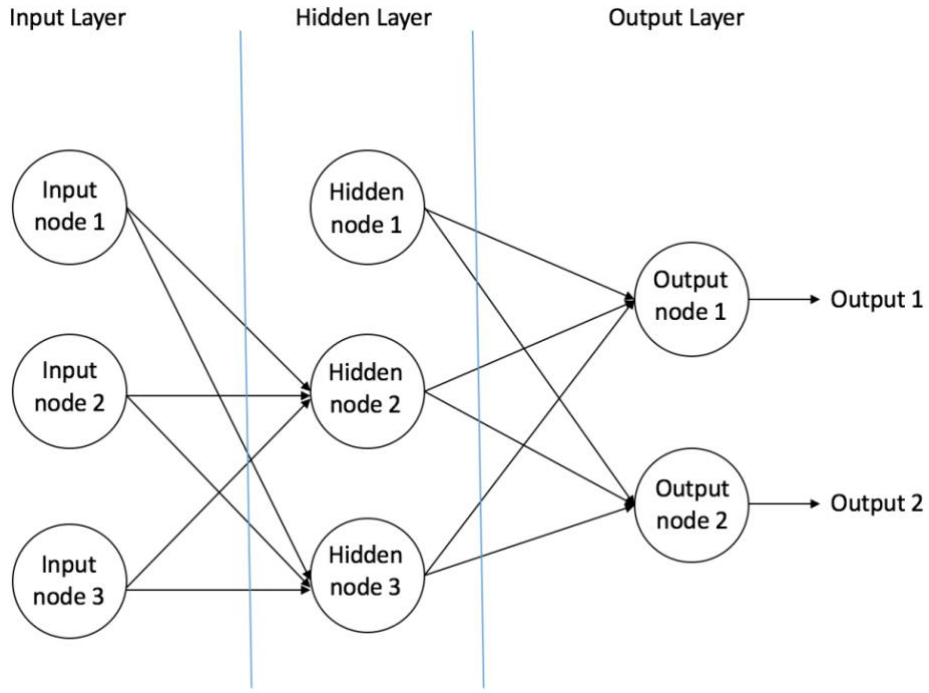


Figure 15: *Illustration of a feedforward neural network containing three layers. Retrieved on 20.02.2021. URL: https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/*

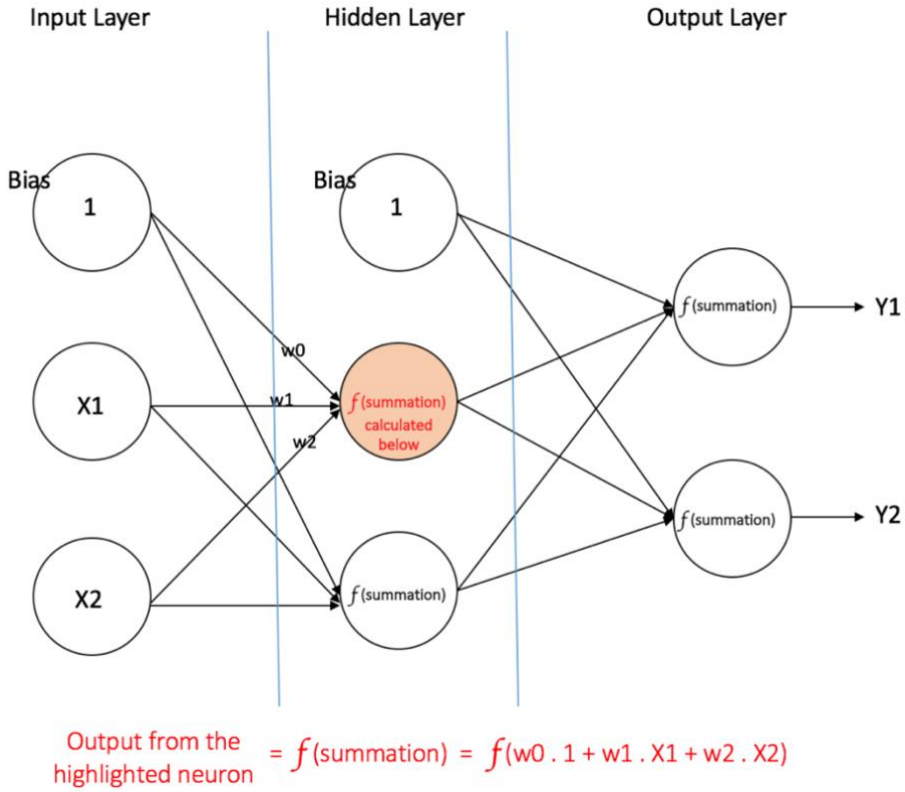


Figure 16: *Illustration of a feedforward neural network containing three layers, including function f in central node. Retrieved on 20.02.2021. URL: https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/*

Now back to the bias. If looking at Figure 16, one can see that all except the output layer has bias nodes (here the weight is 1). Bias in this context means a randomly assigned value taken into consideration when calculating the output. This step is important because it is central in the training part of machine learning. In the given examples of feedforward neural networks, the functions would always stay the same, not adjusting to feedback ("learning"). One process by which MLPs can learn, however, is through a process called *backward propagation of error*, or *backprop* for short. This, put simply, means that the algorithm is learning by mistakes, always self-correcting in order to achieve the lowest possible error rate. This involves adjusting the weights.[21] As one blogger puts it, "the goal of learning is to assign correct weights for these edges. Given an input vector, these weights determine what the output vector is" (Karn 2016b). In supervised learning, this is achieved when the desired output is already known – such as in ImageNet, where the labels are linked to certain images by Turkers before training the algorithm. The image-label combinations are thus treated as a guideline for what is deemed to be the correct answer. A training process is completed when the output error rate reaches a predetermined threshold.[22]

As the initial biases are usually assigned randomly and can thus be quite off at the beginning, it can be hard to parameterize through backprop. This is where so-called deep learning comes in handy. This can, as GitHub user Pouya Sanooei describes it, be thought of as "clever 'tricks' or algorithms that can help us with the training of such 'deep' neural network structures" (2017). Although deep learning can have many forms, the most commonly used for image processing are CNNs, which can be thought of as "add-ons" to MLPs. As output is only as good as its input data, it is crucial for a good algorithm not to drown in "noisy" data, but to single out the most important features in an image to pass on. This is especially important in multilayer neural networks because of the so-called *vanishing gradient*, which is a result of the increasing abstraction from the original data that comes with each hidden layer processing them. CNNs are meant to reduce this problem and can thus be thought of as a kind of automated feature engineering, or "feature detectors" which help reduce the noise and thus enhances the learning in MLPs (Sanooei 2017). In the context of image classification, the CNN involves

[21] For a more thorough, but very simple introduction to this process, I recommend looking up *A Quick Introduction to Neural Networks* on the data science blog. The presentation given by me is also loosely based on this text. URL: https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/.

[22] The predetermined threshold is the same as the acceptable error rate. In light of the discussion in Chapter 1, the process of deciding this threshold equals determining the acceptable level of "fairness." How far from 0 can the error rate be before it is unfair?

adding “receptive fields”, which can be thought of as windows “sliding” over the image. As Sanooei describes this process concisely, I quote a paragraph here in full:

> We then connect those "receptive fields" (for example of the size of 5x5 pixel) with 1 unit in the next layer, this is the so-called "feature map." After this mapping, we have constructed a so-called convolutional layer. Note that our feature detectors are basically replicates of one another – they share the same weights. The idea is that if a feature detector is useful in one part of the image it is likely that it is useful somewhere else, but at the same time it allows each patch of image to be represented in several ways. Next, we have a "pooling" layer, where we reduce neighboring features from our feature map into single units (by taking the max feature or by averaging them, for example). We do this over many rounds and eventually arrive at an almost scale invariant representation of our image (the exact term is "equivariant"). This is very powerful since we can detect objects in an image no matter where they are located (2017).

In short, this means that the CNN works as a kind of filter through which an image is processed by the algorithm, extracting the features deemed most important via the pooling layers, reducing the vanishing gradient considerably.

To really understand these processes, it is pertinent to keep in mind that for any object recognition system to work, the image must first be broken down and represented as a matrix of pixel values given as numbers. In a black and white photograph, this could mean that the value of black is 0 while white is represented as 1, with all other values distributed on a scale between these limits. When a frame within the image is analyzed, various features can be detected as contrasts in number values. In CNNs, it is possible to apply a number of filters for analyzing the image, which are sensitive to different features.[23] Different *feature maps* will be produced depending on which filter is applied. These are, easily put, versions of the same image, but where only what is “caught” by the specific filter has been left visible. Next, the layer containing feature maps is submitted to a process called *pooling*, which in short means that every section of each feature map is broken down to contain only the most important information, thereby reducing the dimensionality of the feature map considerably. Together, one round of feature maps and pooling is referred to as a *convolution*, and it is up to the programmer how many convolutions are performed on the images by the algorithm in the recognition process. The pooling process is thus repeated a given number of times until the information arrives at the output layer.

[23] This is only a very general description of the process, leaving out the more technical mathematical explanations. For a more detailed introduction to CNNs, I recommend *An Intuitive Explanation of Convolutional Neural Networks* by user “ujjwalkarn” on the data science blog. URL: https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/.

Although the process as I have presented it here is greatly simplified, it is clear that the final output is based on a series of abstractions which are in turn based on numerical representations of the original image (Fig. 18). The outermost layer, however, is usually a version of an MLP, as described above. The final step, performed in the output layer, is to summarize the information passed on through the convolutional layers, and to classify this input according to the classes it knows from the training dataset. The final output can then be represented either as a label alone, or along with an estimated probability of whether the label is correct.[24] Based on the above description, it should not be surprising that the abstract shapes in *How do you see me?* could come to be misrecognized as faces by the program. On a general level, the convolutional process is a series of feature extractions, and looking at Dewey-Hagborg's images we get an idea of what the algorithm understands as her face's important features. While it is the programmer who decides parameters such as number of filters and filter size, it is the algorithm itself that comes up with the filter values through the training process.

This takes us to *gradient descent*, which is the concept or model guiding most training processes (Fig. 17). The principle is often described in terms of a landscape where one is trying to arrive at the lowest possible altitude (error rate), but has no map with which to plan the route (Tørresen 2013: 30). By learning according to the gradient descent principle, the algorithm has to sort out which "road" will lead to the bottom of the valley, or *global minimum*, rather than simply a local "dip" in the terrain. As on a real hike, the steepest slope does not necessarily lead to the lowest overall-altitude, and the process might therefore include several direction changes along the way. When the algorithm finally decides on a certain path, there is per now no way for the programmer to know how this particular route was decided upon, or why it supposedly works better than the alternatives. Gradient descent is therefore central to the opacity of the training process in machine learning, illustrating well the field's general reproducibility problem: as with alchemy, one can arrive at the desired results without really knowing how.

It is not surprising the, that when the kind of abstractions described above were paired with an automated evolutionary process, the AFR system used in *How do you see me?* could arrive at something that makes so little sense to human eyes. Admittedly, a programmer would probably not think of applying the evolutionary method in the way Dewey-Hagborg did, as doing so serves no practical or commercial goal. The project can thus perhaps be understood best as a kind of thought experiment. As is argued by Ben Recht, a computer scientist at the

[24] This is what happened in the installation version of ImageNet Roulette discussed in Chapter 1, where the software guessed my age, gender and emotional state with decimal numbers representing probability in parentheses.

University of California, Berkeley who coauthored the alchemy talk discussed in Chapter 1, AI needs to borrow from the field of physics, which is adept at devising experiments to root out explanations for phenomena (Hutson 2018). In this sense, *How do you see me?* serves a crucial role in functioning as a model to visualize what defines a face to this particular facial recognition system. In asking what and how the machine vision "sees", Dewey-Hagborg's project makes it apparent that it makes little sense to think of image recognition in terms of human intelligence, but rather as a sort of pattern recognition and probability calculation model. It is thus clear that the notions of scientific objectivity often ascribed to artificial intelligence is unjustified. This being said, I still wonder what this means for notions of intelligence in computers. Since these are the terms with which image processing systems are most often described, such questions must be discussed if we are to have a good understanding of the kind of systems that are arguably the most important curators of images today.

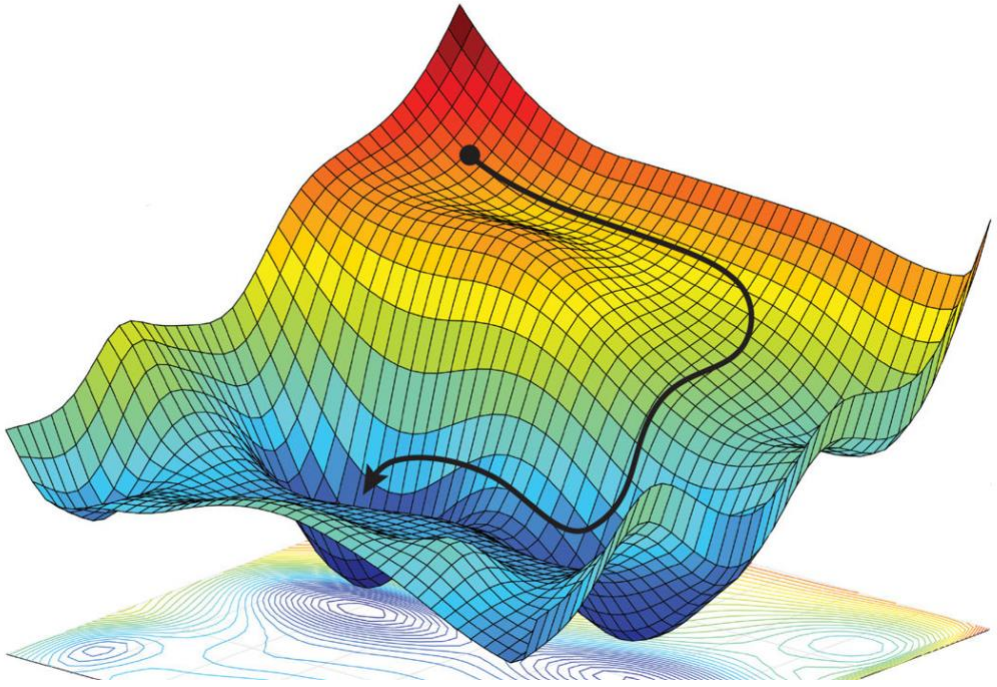


Figure 17: *Illustration of the trial-and-error method referred to as gradient descent. By Alexander Amini and Daniela Rus. Massachusetts Institute of Technology, adapted by M. Atarod/Science. Retrieved on 20.02.2021. URL: https://www.sciencemag.org/news/2018/05/ai-researchers-allege-machine-learning-alchemy*

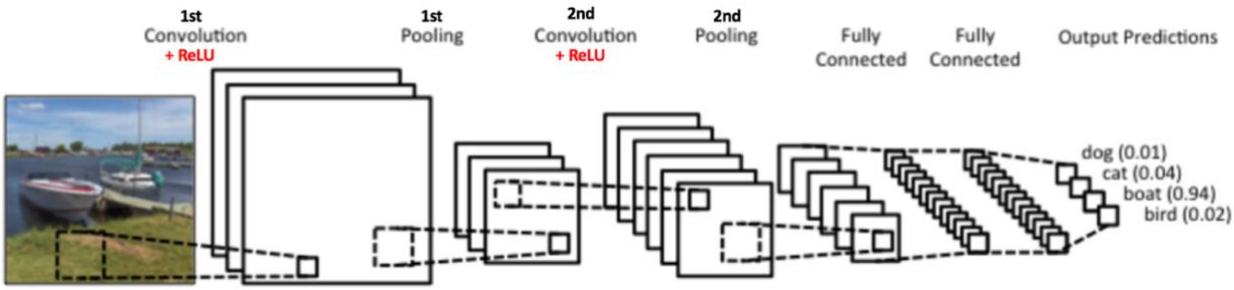


Figure 18: *Illustration of the process involved in image recognition in CNNs. The first input here is an image of a boat, which is broken down into three separate feature maps, which are in turn exposed to a pooling process. In the second layer, the process is repeated before the information arrives in the fully connected layer (meaning that all nodes correspond directly to another one). Finally, the MLPs in the fully connected layer produces an output, assigning a label along with probabilities. Retrieved on 21.02.2021. URL: https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/*

A Brief History of AI

Whether machines can be intelligent is a question at the center of AI research. For centuries, Western thought had been based on the Cartesian principle of mind and body (or mechanism and intelligence) as separate entities. It was not before the 1900s, however, that mathematician started to discuss seriously whether it was possible to *axiomatize* mathematics: the goal was to find a set of purely logical principles that could be arrived at in a step-by-step procedure, as if mechanized (Rees 2020: 112). In his 1936 article, *On Computable Numbers*, the British mathematician Alan Turing (1912-1954) proposed to take this idea literally, arguing that if such principles could in fact be developed, then a digital computer[25] should theoretically be able to perform the process. He then proposed that such a thinking machine could indeed be built if the process was broken down into a finite series of steps – just like a human computer solving equations with a rule book (Turing 1937 [1936]).

While Turing at the point of publication probably only imagined this as a theoretical possibility, it is widely known that the Second World War brought about the need for machines that could decode encrypted messages automatically – an historic effort in which Turing's role in developing the fabled Enigma machine earned him a place on the £50 note. After the war, the stage was thus set to bring this previously hypothetical line of research into practice. In the widely influential article *Computing Machinery and Intelligence*, Turing described a set of research areas that he found especially promising as paths towards "true" AI, proposing that digital computers of the future could be able to perform a series of complex operations in the

[25] The term then simply referred to a calculating human, rendering a *digital* computer a revolutionary idea.

"purely intellectual fields." According to him, the most promising such activities were "the playing of chess", as well as "understand[ing] and speak[ing] English." Furthermore, computers should be programmed to "follow the normal teaching of a child. Things would be pointed out and named, etc." (Turing 1950: 460). As will be made clear, these proposals would come to influence the future of AI research up until this day. Through his research, Turing not only contested the idea that machinery and intelligence were exclusive concepts; he also proposed that the reason intelligent machines had not yet been built was because they had hitherto only been imagined as "slaves" to serve specific, limited tasks (Rees 2020: 196). This can probably be seen as a consequence of the predominance of Cartesian dualism which rendered thinking machines oxymoronic.

With help from some other notable figures, such as John McCarthy (1927-2011), the stage was set by the early 1950s for a new field of research trying to produce thinking machines. A pivotal point in time that is referred to in virtually all histories of AI, is the Dartmouth Summer Research project on Artificial Intelligence in 1956: an inaugural event bringing a younger generation of researchers together, and famously where the term *artificial intelligence* was coined. The philosopher Tobias Rees has argued that the reason behind this specific phrasing was to distance the emerging generation from the effort to build "electrical brains," an idea which had widely influenced the older researchers. The newcomers, on their side, turned to behavioral psychology, arguing that the field's major obstacle was not a problem of computing capacity, but rather a failure to create suitable programs for the machines they had available (Rees 2020: 114).

Based on a strictly behavioristic understanding of intelligence, their main goal thus became to write software, or code, that modeled the cognitive problem solving of a human in a way that could be implemented and run on a computer: the idea was that if a computer could do what a human could, then it must be considered intelligent. Research thus focused on "translating" cognitive processes into symbolic language. This led to the first information-processing language, IPL-I, created by Newell and Simon at the Carnegie Institute of Technology in 1956. This language was, as Rees tells it, based on the idea that "a human's knowledge was roughly equivalent to storing symbolic representations of things in the world, and intelligence was roughly equivalent to the algorithmic search through this stored information and, where necessary, the manipulation of these symbols" (Rees 2020: 114).

During the next decade, many alternative information-processing languages were made, but they were all based on this shared assumption which equivalated knowledge with stored symbolic representation and intelligence with algorithmic search through, or manipulation of,

such symbols. When I now continue to Hubert L. Dreyfus' critique of AI published in 1972, it should be noted that it corresponded to the situation as told up until this point. I will discuss this now before returning to subsequent developments in artificial intelligence later.

Intelligence as Situated Knowledge

The general argument in *What Computers Can't do: A Critique of Artificial Reason*[26] can be broken down to three main parts. Firstly, Dreyfus argues that one of the reasons that AI research at the time of writing had not reached the sophistication hoped for since Turing, was that digital computers are per definition not capable of *semantic understanding*. He explains how early language processing programs faced some difficulties regarding understanding, as opposed to purely technical processing. His critique is mainly focused on the program STUDENT, which was in the 1960s considered the most advanced language program. Even so, it was limited to processing sentences that could be formulated mathematically. Daniel Bobrow, its developer, was open about this setback, noting that when he used the word *understanding*, what he really meant was an operational definition of what computers do when processing sentences that fit exactly inside the parameters of the program: "a computer *understands* a subset of English if it accepts input sentences which are members of this subset, and answers questions based on information contained in the input. The STUDENT system understands English in this sense" (as quoted in Dreyfus 1992: 134). Dreyfus' critique to a large extent continues along these lines, presenting a series of examples where computer programs clearly lack the abilities that we would normally attribute to semantic understanding if speaking of human beings. He also points to how the STUDENT program was limited to only a small amount of the English language.

In line with what I have indicated in relation to images in this and the previous chapter, Dreyfus also takes care to note that what computers do is simply to break down simple sentences according to a pre-given set of rules. In this respect, he finds it extremely problematic that Marvin Minsky, who was at that time a professor at MIT – and perhaps the most influential person in the history of AI after Turing[27] – held that STUDENT demonstrated that "understanding machines can be built" (Dreyfus 1992: 134). True, Minsky did admit that the program had room for improvements, but he was nevertheless guilty of what Dreyfus calls *the*

[26] The book has been published in several versions: the first, *What Computers Can't Do*, was published in 1972. Dreyfus later published a revised version named *What Computers Still Can't Do* in 1979, and again in 1992. I am here using the newest, 1992 version.

[27] In the book *Artificial Unintelligence*, Meredith Broussard writes: "Look behind the scenes at the creation of virtually any high-profile tech project between 1945 and 2016, and you'll find Minsky (or his work) somewhere in the cast of characters" (2018: 69).

first step fallacy. This implies assuming that simply because the first step towards a goal has been accomplished, complete success is guaranteed in the long run (Dreyfus 1992: 136). Dreyfus also questions the underlying assumption on which these systems are based, namely that algorithmic information processing has anything in common with how humans process language – or anything other than mathematical problems: "why", he asks, "do those pursuing Cognitive Simulation assume that the information processes of a computer reveal the hidden information processes of a human being, and why do those who work in artificial intelligence assume that there must be a digital way to perform human tasks?" These questions lead him to identify and contest four general assumptions informing AI research.

Dreyfus starts this part of his examination with what reads almost as a sigh of despair over what he perceives as unfounded optimism towards AI. The one fallacy that he is most focused on, is the consistent belief across the field that human intelligent behavior is somehow the result of information processing by a digital computer (the brain), and that human and digital information processing can therefore be equated. As an example of this, he quotes a research paper by Newell and Simon (the duo behind IPL-I discussed above):

> It can be seen that this approach makes no assumption that the "hardware" of computers and brains are similar, beyond the assumptions that both are general purpose symbol-manipulating devices, and that the computer can be programmed to execute *elementary information processes* functionally quite like those executed by the brain (1992: 155).

But this, Dreyfus notes, is no "innocent and empty assumption" (1992: 155). On the contrary, both the concepts of a "general purpose symbol-manipulating device" and "elementary information processes" are just as indefinite as they are ungrounded. Furthermore, the assumptions implicit in this statement are revealing of an overall ad hoc approach to human intelligence within the field of AI: because the goal is to create digital computers capable of replicating human intelligence and the only kind of machines that have been built are general purpose symbol-manipulating devices requiring information to be presented as discrete elements, then this must be how humans think. Dreyfus breaks this overarching assumption further down into four, which he terms 1) the biological assumption, 2) the psychological assumption, 3) the epistemological assumption, and 4) the ontological assumption.

The *biological assumption* supposes that at some level, the brain processes information in discrete operations by way of a biological equivalent of on/off switches in neurons. While this is certainly a simplification of the neural system (and the brain), it is not, in my understanding, too far off from present-day medical consensus. I therefore leave the biological

assumption at that. The *psychological assumption*, however, is the one outlined in the previous paragraph, supposing that the mind can be viewed as a kind of device operating on bits of information according to formal rules. As Dreyfus argues, this idea can be traced back to both enlightenment empiricists and idealist (notably Hume and Kant), who "prepared the ground for this model of thinking as data processing – a third-person process in which the involvement of the 'processor' plays no essential role" (1992: 156)." This is in turn related to the *epistemological assumption*, proposing that all kinds of knowledge can essentially be understood and expressed in terms of logical relations, or Boolean functions.[28] In this sense, every kind of information, no matter what, is in principle translatable to algorithmic language and processable by computers. When Minsky and Turing propose that humans can be understood in the same way as a Turing machine, Dreyfus thus takes this to mean the following: "a digital computer can reproduce human behavior, not by solving physical equations but by *processing data representing facts about the world using logical operations* that can be reduced to matching, classifying and Boolean operations" (1992: 196).

Finally, the *ontological assumption* is based on the idea that because all information fed into digital computers must be in bits, "the computer model of the mind presupposes that all relevant information about the world, everything essential to the production of intelligent behavior, must in principle be analyzable as a set of situation free elements" (Dreyfus 1992: 156). The problem with these assumptions, Dreyfus argues, is not the idea that things can be represented in a certain way or described as a kind of digital formalism in itself. The problem is that researchers such as Minsky seem to confuse this fact with the idea that their representations are somehow literal translations of human processes without loss of complexity. A problem with this is that because computers are not themselves involved in the situations they analyze, every bit of data always has the same value (1992: 201). In this sense, computers are not what Kant would call "transcendentally stupid," seeing how they are at least capable of applying rules to a specific case unambiguously described as rules beforehand. They are, however, what Dreyfus calls "existentially stupid," meaning that computers cannot cope with *situations*: "they cannot accept ambiguity and the breaking of rules until the rules for dealing with the deviations have been so completely specified that the ambiguity disappears" (1992: 201). In other words, computers, unlike humans, are unable to simply "go with the flow," to assess and feel their way forward in a reciprocal relationship to the things and people around

[28] Boolean algebra is a branch of algebra in which all variables are either true or false. Values are usually denoted as 1 or 0.

them. If this problem is to be overcome, Dreyfus says, "AI workers would have to develop an a-temporal, nonlocal theory of ongoing, situated, human activity" (1992: 201). In other words, a highly improbable task.

What the psychological, epistemological and ontological assumptions have in common, is that they all assume that humans are devices; always calculating according to rules on data in the form of atomic facts. This view, Dreyfus argues, can be understood as a confluence between two powerful philosophical streams:

> First, the Platonic reduction of all reasoning to explicit rules and the world to atomic facts to which alone such rules could be applied without the risk of interpretation; second the invention of the digital computer, a general-purpose information processing device, which calculates according to explicit rules and takes in data in terms of atomic elements logically independent of one another. In some other culture, the digital computer would most likely have seemed an unpromising model for the creation of artificial reason, but in our tradition the computer seems to be the very paradigm of logical intelligence, merely awaiting the proper program to accede to man's essential attribute of rationality" (1992: 235).

Thus, the idea of intelligence as imagined by the AI industry did not appear out of nothing but is rather the result of a long tradition within the Western philosophical canon. The problem at hand is thus far more complex than simply to argue against the possibility of artificial intelligence per se. In light of this, Dreyfus' finds himself faced with the challenge of proposing an alternative idea of human intelligence. His main point, building on the phenomenological tradition after Husserl, is that the *body* plays an important role in intelligent behavior; that the "sense of overall context may organize and direct our perception of details" (1992: 241). This forms a stark contrast to computers, which must either take into consideration every single bit of information presented to it, or none at all.

Although a human perceiver, like a machine, also needs feedback in order to find out if he or she has successfully recognized an object (if that is the task), Dreyfus points to an important difference in the feedback involved: while a machine can, at best, make a specific set of hypotheses and then find out if they have been confirmed or refuted by the data, a human "can constantly modify its expectations in terms of more flexible criterion" (1992: 250). In his view, the question of whether artificial intelligence is at all possible therefore boils down to the question of whether there can be an artificial embodied agent. While this question, if AI is possible, goes beyond the scope of this thesis, I nevertheless find Dreyfus' critique relevant to the discussion of *How do you see me?*. Especially because it situates the problems facing AFR within a history of ideas, demonstrating how certain conceptions about computers came about. But how, I wonder, does the above discussion read in light of later developments?

Fast Forward: Image Recognition in 2019

To pick up where I left off in the history of AI, things have changed drastically since the 1960s. In the case of object detection- and recognition in images, the techniques have been greatly improved by the availability of online images since the 1990s. What's more, speeds and memory capacities of computers have also been greatly enhanced by material advancements such as silicone chips. It was also not until the early 2000s that deep neural networks became available. As Tobias Rees tells it, the changes that started to develop around this point were partly due to a young generation of engineers who came up with a new method of building neural networks (2020: 115-16). What is today known as *deep learning* was developed by Geoffrey Hinton in conversation with Yann LeCun and Yoshua Bengio. The name refers to the fact that the more data the network is fed, the better it becomes at the prescribed task. ImageNet also played a central role in these developments, in that the repository of tagged images it provided set the stage for experimentation and refinement of different neural network architectures and algorithms. Altogether, this has led to a massive improvement in performance, new systems outperforming older models significantly in all fields. These developments were further boosted by the arrival of venture capitalists in tech companies, which led to an influx in data, computing power and financing, and even quicker advancements (Rees 2020: 115-16).

Since the 1970s, performance in pattern discrimination has been greatly enhanced. Most tech companies also seem to have stopped referring to this specific field as AI, in favor of terms such as machine learning, or even the more technical sounding term *logistic regression*. Nevertheless, projects such as *How do you see me?* make it clear that Dreyfus' main points are still relevant today. This view is in opposition to Tobias Rees' who points to how the early AI – even if undermining the intelligence-mechanism dualism – still was based on the idea of human exceptionalism in trying to develop programs mimicking human intellectual processes. Contrary to this, he proposes that intelligence from a deep learning perspective is not an exclusive human property. On the contrary, Rees argues, intelligence *is* a property of neural networks (2020: 116). While it is not my interest to revoke the idea of human exceptionalism regarding intelligence, I want to point out that the idea of intelligence that Rees here proposes in relation to artificial neural networks is at best a very limited one. Even if the processes that can be performed are regarded as a kind of intelligence simply because of their speed and scale, it is evident by the analysis in this chapter that neural networks are still "existentially stupid" in their limitation to unambiguously described tasks. So, even if granting a relativized definition, this does not actually affect the notion of humans as unique beings. This also renders the term *recognition* in image recognition dubious, as it plays on human perception. In light of

the above discussion, it is clear that speaking of these processes in terms of human understanding is to anthropomorphize the technology.

Another crucial difference to human perception of the world is that even the most advanced CNNs are limited to representations of the world *through images*. While this is not a problem in itself, it is nevertheless problematic precisely because the algorithm as such has no understanding of this difference (and it seems that neither do some programmers). Moreover, just like Dreyfus pointed out how Turing and Minsky confused their own programs as non-reductive translations of the world without loss of complexity, it seems like the situation today is the same in machine learning: the abstractions are confused as one-to-one representations of reality. Today's CNNs are also reliant on access to a dataset of carefully labeled images in order to perform self-improvement through backwards propagation, a process which I in the previous chapter demonstrated to be full of bias. While bias is certainly not limited to computers (the labeling process by Turkers makes this clear), I argue that this is even more problematic in machines precisely because they – as opposed to humans – lack the ability to "say no."

This takes us back to Sontag's reflection on understanding as rooted in the ability to reject what is presented in a photograph. Algorithms clearly have no such ability: it can only improve itself in the sense that it can statistically develop tactics that let it arrive more effectively at the labels provided to it, whose correctness is predetermined by whoever created the training database. At no point in this process does the algorithm have the possibility to question the implicit assumptions in the dataset, or the parameters to which it is programmed to adhere. Never will or can it ask questions like "*is this face representative of all faces in general*?" or "*is this image of two small children really a good representation of the word bastard*?" Rather, it must take all the data at face value. Even if accepting that this is indeed some kind of limited intelligence, it is therefore clear that we are not here faced with anything close to what is usually referred to as understanding.

Yet another difference between human perception and image recognition algorithms that I would like to point to, is that humans are capable of seeing more than one thing in the same image – as exemplified by the classic duck-rabbit illusion. Reflecting on this, Lila Lee-Morrison, in her thesis on AFR, contrasts Francis Galton's (1822-1911) composite portraits to those of Ludwig Wittgenstein (1889-1951). Galton's goal was to produce composites in order to arrive at a statistical average so that he could produce a generalization that would let him depict the features of a supposed "criminal type." Wittgenstein, on his side, saw something else entirely in his pictures. Where Galton saw the composite as representing *one* ideal form, Wittgenstein, as Lee-Morrison argues, approached it as an image of a *thought process*. This let

him connect it to his concept of family resemblance. As she describes it, Wittgenstein argued against Galton's theory of abstraction derived from John Locke, claiming that knowledge derived from a logic of reduction is *antithetical to the work of a philosopher*. The tendency of mathematicians to reduce topics to generalizations, he writes, leads the philosopher into *complete darkness* (Lee-Morrison 2019: 101). Contrary to Galton, when Wittgenstein looked at his own composites made from images of him and his three sisters, it presented him with a perceptual enigma: on the one hand, he could see a single face, but on the other, he could simultaneously see every sibling's individual features. In this, Lee-Morrison writes, Wittgenstein was able to see the composite as "an image of multiple perceptual outcomes rather than a *singular* probability of a type" (2019: 107).

This last point is crucial if one is to understand image recognition in terms of perception. As has already been discussed, image recognition systems are programmed to produce output in the form of a probability between 0 and 1. What I have not mentioned until now is that they are also usually programmed so that the sum total of all probabilities cannot exceed 1. This is ensured by using something called a Softmax function in the output layer, which takes a vector of arbitrary real-valued scores and squashes it to a vector of values that sum to one (Karn 2016a). If the input image is a boat, the target probability would then be 1 for the category "boat", while all other classes would get a score of zero. This makes sense in terms of probability, which can, as most are probably aware, not exceed 100 percent: as image recognition systems are essentially a kind of probability calculator, the goal when programming them is that *one* object should only be labeled as *one* thing and that thing only. But who decides what that one thing is?

Going back to the boat-example, it is clear that any such picture might indeed represent a number of things: is it a fishing boat, a family boat, or perhaps even a toy boat? And what if it is not the boat itself that I am interested in? What if, for me, the image represented the feeling of summer, fun, vacation, or even a happy family time? Or perhaps, as Roland Barthes put it in *Camera Lucida*, a photograph represents *death* – a moment no longer in existence. These problems are apparently lost to anyone who talks or writes of object recognition as objective tools. Whether we refer to this in terms of intelligence is not important. What matters is that we are here faced with a kind of technology that reduces the human experience of the world into computable information through a series of complicated abstractions while neglecting to acknowledge the loss of complexity that this constitutes. The real problem thus lies in the fact that the abstractions are treated as if they were the "real thing."

In light of the above, the problems described in this chapter seem to recapitulate the message in René Magritte's famous *The Treachery of Images* (1928-29); a painting of a pipe with the caption "Ceci n'est pas une pipe" (this is not a pipe). As is by now a commonplace observation, the painter's point was that the painting was but a *representation* of a pipe. That this difference is apparently lost to image recognition systems has also been noted in a not-yet mentioned work by Trevor Paglen. This image, named *The Treachery of Object Recognition* (2019), is a version of Magritte's work where the pipe is replaced by an apple (Fig. 19). With the help of an algorithm (also trained in ImageNet), Trevor Paglen let a digital version of this painting be categorized. Tellingly, in the version presented to the audience, four green squares and labels have appeared: "a large white sign," "black and white sign," "a large green leaf," and "red and green apple." In this way, the apple in Magritte's painting works as a metaphor for the entire discussion presented in this chapter, illustrating precisely how the algorithm is unable to differentiate between an actual apple and a painting of one.

As Paglen puts it in an exhibition catalogue, this also points to a certain politics ingrained in the ability to name something: "when Magritte says 'this is not an apple,' I think of that as a microcosm for a kind of self-determination. It's akin to civil-rights activists saying 'I am a man'" (Pardo 2019: 44). In this sense, the juxtaposition of Magritte's work with an object recognition algorithm makes it clear that the technology is, through its inflexible machinic perception, robbing people of the possibility of self-representation. While this has always been inherent to AI, its consequences today are much greater as automated systems are larger, faster, and integrated in ever more aspects of life.

Conclusion

Dewey-Hagborg's project demonstrates that when set to test under abnormal circumstances, an AFR system can misidentify seemingly random shapes as faces. This has serious consequences for such systems' credibility when applied to real world situations. Through a reading of *How do you see me?* this chapter has problematized some presumptions inherent in image recognition systems. Most importantly, the neural network architectures discussed seem to presuppose an ontological connection between the representations processed by the system and the objects themselves because the algorithms are engineered as if objects exist in a world devoid of situation. I have shown that these assumptions can be traced back to the early history of AI and the question of whether it is possible to axiomatize mathematics. When this idea was first transferred to the area of digital computers via Turing, the practical circumstances limited their use to simple algorithmic problem solving. As Hubert L. Dreyfus argued extensively in

What Computers Can't Do, this in time led to a general confusion within the field, equating the process of algorithmic problem solving with how humans think in general. Dreyfus also showed how these ideas were not at all new to Western thought, which is why he supposes that the ideas he criticizes could pass as "neutral" for so long. The connection between AI and Enlightenment ideas will be made even clearer in the next two chapters.

Although the field of image recognition has made some amazing technical progressions since its early days, this chapter has demonstrated that the four underlying assumptions that Dreyfus pointed to in 1972 are still implicit in the Convoluted Neural Networks of today. While some philosophers, like Tobias Rees, propose that CNNs constitute a rejection of human exceptionalism, I have been critical towards this idea. My goal, however, has not been to discuss the possibility of artificial intelligence per se, but rather to analyze the manner in which machinic ways of "seeing" and "thinking" translate to normal conceptions of these terms. In answer to Dewey-Hagborg's question, then, it seems that the algorithm does not "see" her at all, but rather processes representations of her to calculate the probability of the image being the artist.

The discussion makes it clear that if CNNs are said to be intelligent, it is only in name and not in practice. Other more precise terms are thus welcome. A critical factor to this conclusion is the fact that at no point do the systems have the opportunity to question or contest the taxonomies informing their training process. Neither do neural networks have the capacity to differentiate between objects and the representations thereof via images. I, like Dreyfus, do not believe mind and body to be separable entities. It is evident through the analysis of *How do you see me?* that the advent of CNNs do not give credence to the idea that machines can think. On the contrary, their inflexibility demonstrates exactly how important embodied situatedness is to human understanding. Notions of intelligence or understanding in relation to image recognition systems should therefore be applied with great care. As an alternative, I propose to talk of them as what they truly are, which is algorithmic pattern recognition systems whose goal it is to produce singular outputs in terms of probability. Only if we see automated object recognition in this way can we begin to understand the nature of contemporary networked digital photography.

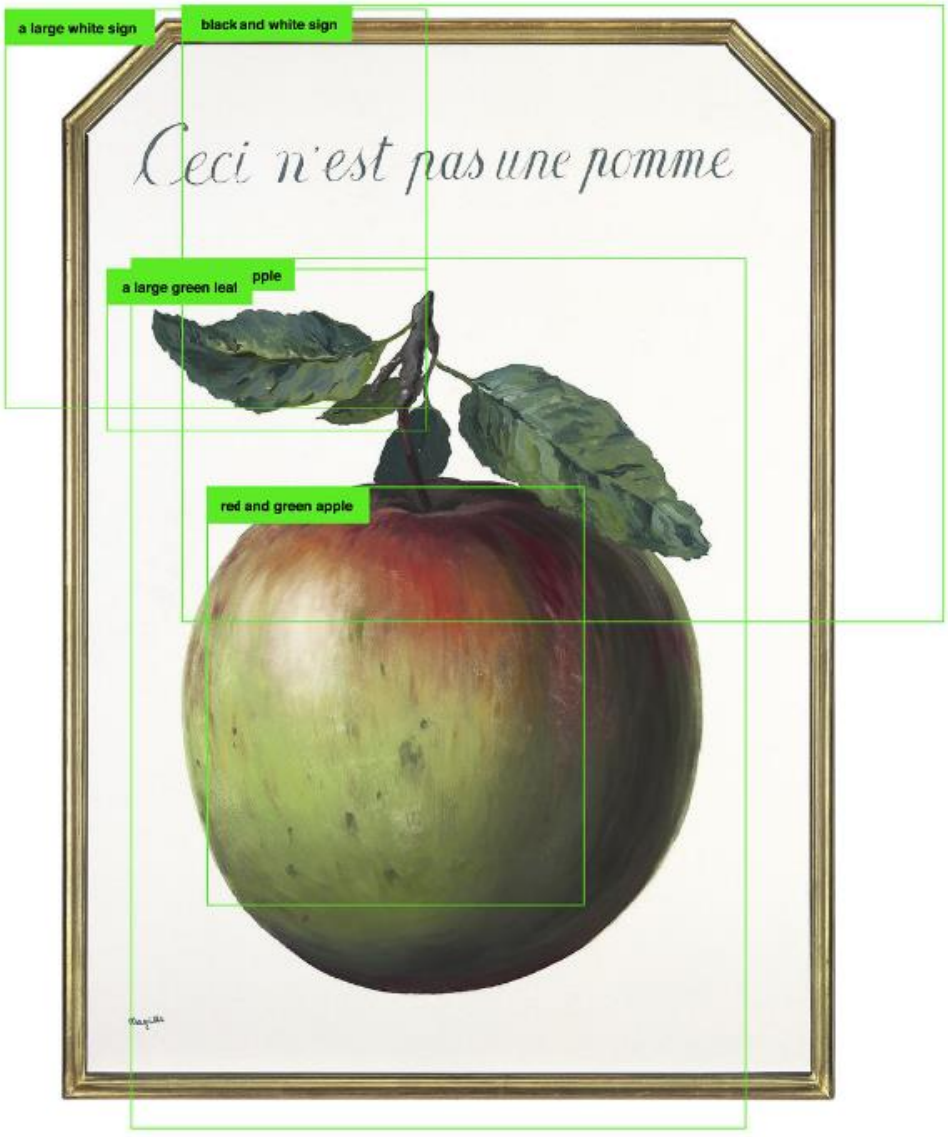


Figure 19: The Treachery of Image Recognition *(2019) by Trevor Paglen. Dye sublimation metal print, 39,4 x 33 cm. Image retrieved on 25.02.2021 from Artsy. URL: https://www.artsy.net/artwork/trevor-paglen-the-treachery-of-object-recognition*

Chapter 3

A History of Social Sorting: *Spirit is a Bone* by Broomberg & Chanarin

Introduction

This chapter is concerned with a series of photographic portraits which I first encountered on a vacation in Hamburg in 2018. I am here referring to the work *Spirit is a Bone* (2013-15) by the artist duo Adam Broomberg (b. 1970) and Oliver Chanarin (b. 1971), or "Broomberg & Chanarin" as they called themselves before splitting up in February 2021 (Buck 2021). Coming from South Africa, the two photographers both live and work between Berlin and London, teaching photography on the side of their artistic practice. While the duo's early projects were examples of relatively classical documentary photography, they gradually developed strategies increasingly critical to the idea of photography as an objective source of documentary truth. If there is any common thread amongst the works from their two-decade long partnership, it is perhaps that they seek to scrutinize power relations – especially those in which the medium of photography is itself involved in serving power structures.

Spirit is a Bone is no exception to this rule. Analogous to ImageNet Roulette, this project is a scrutiny of the practices of categorization involved in facial recognition technology. But where the projects discussed in the two previous chapters have a predominantly technical approach to the subject matter, *Spirit is a Bone* is more focused on the technology's ideological basis, explicitly situating the portraits within an art historical context. The work exists in many versions, but in order to limit the scope of the following analysis, I will focus on the book version as well as the installation presented in the exhibition *[CONTROL] NO CONTROL* at Hamburger Kunsthalle in 2018. While a presentation of this exhibition as a whole would surely be interesting, this is besides the scope of this chapter. I will therefore not include a survey of the installation in relation to the other exhibited works. Instead, I will focus on the aesthetic encounter with the installation in the gallery space on its own terms before moving on to a presentation of the book.

Regarding structure, I have chosen to present the project in the order in which I personally encountered the two versions, as this enables me to retell my experience as it unfolded. After presenting these encounters, I move on to discuss some deeper issues alluded to in the book. An important aspect of *Spirit is a Bone* is how it appropriates the image universe

of the 20th century portrait photographer August Sander (1876-1964). By categorizing the portraits according to his schematic, Broomberg & Chanarin raise questions concerning the practice of this particular photographer while also connecting their own work to a history of disciplines concerned with social sorting going all the way back to physiology, craniometry and criminology in the 1800s. This discussion not only offers an interesting perspective on the history of photography, but also invites a reading of modern biometrics as part of a long history of positivist identification schemes aimed at "looking past" the deceptive layers of the face. As the portraits were made with a contemporary AFR system, Vocord 3D, *Spirit is a Bone* offers an interesting entry point to this discussion in that the technology parallels craniometry and phrenology in having the human skull as its object of study. Building on the essay *The Bone Cannot Lie* from the book version of *Spirit is a Bone*, I will finally discuss whether there are some ideological similarities between those historic systems of categorization and present-day technology.

First encounter: *Spirit is a Bone* as Installation

It was an exhaustingly warm summer day in Hamburg. After hours of sightseeing, it was thus a welcome relief to finally enter the delightfully airconditioned premises of Hamburger Kunsthalle. As I had not actually planned on going there, I was positively surprised when it became apparent that the museum, as part of the city's ongoing Triennial of Photography, was currently displaying an exhibition concerned with surveillance.[29] Excitedly, I started exploring the exhibition. Turning a corner, something caught my eye in the room ahead of me: a large collection of faces in black and white detached from their bodies on a pitch-black background. These uncanny faces immediately ignited my curiosity, drawing me towards them. As I got closer, it became clear that what from a distance had looked like one large image was in reality a series of separate glass plates placed closely together, collectively covering a large wall area (Fig. 20). As I stood there, face to face with the work, my immediate thought was that these portraits appeared strikingly *inhuman*.

The primary reason for this impression was perhaps the assaultive manner in which the faces had been completely detached from the rest of their bodies: like masks, they were cut along the persons' hairlines and contours, revealing no part not strictly belonging to the face.

[29] I am here referring to the exhibition *[CONTROL] NO CONTROL* as presented at Hamburger Kunsthalle 08.06.-23.09.2018. The exhibition was part of the 7th Triennial of Photography Hamburg 2018. For more information on the exhibition, see the triennial catalogue *Breaking Point: Unlearning, Rethinking, Restarting* (Stuttgart: Hartmann Books & Triennale der Photographie Hamburg, 2018).

Contrary to masks, however, the eyes were intact, but the gazes were truly unsettling. As opposed to what I am used to with images of people, looking at these portraits gave me no sense of understanding who the subjects were or their states of mind. Neither did the portraits reveal even the slightest hint about the context in which the shots might have been taken. As exemplified by Michel Foucault's famous analysis of the theatre of gazes in Velázquez' *Las Meninas* (1656), a portrait subjects' eyes can be the source of an ocean of meaning (Foucault 2006 [1966]: 23-40). Facing *Spirit is a Bone*, however, the eyes were strikingly non-communicative – as if the subjects did not even interact with the photographer at the time of caption. On the contrary, the only telling thing about the portraits was the very lack of expressiveness itself. Therefore, the only thing of which I felt quite certain at this point was that these people were not aware of being photographed.

On closer inspection, I also noticed something else: several of the faces had irregularities, or "glitches," indicating some kind of digital process in which the parts had been taken apart and reassembled – some more seamlessly than others. Some of the faces were indeed so twisted that they reminded me of cubist paintings portraying the sitter from various perspectives at once. But contrary to such, these images were clearly situated in some a digital space – a conception substantiated by the medium itself; C-prints on black glass plates. The stark contrast between the bright faces and the black background gave the faces the appearance of hovering or floating in an abstract, immaterial space, which is how many people, despite its flaws, conceive of the Internet and digital phenomena (Gronlund 2017: 88). But the effect of detachment was counteracted by the properties of the prints themselves, in that the glass reflected the work's own immediate surroundings.

Facing the installation, I was thus also facing my own reflected likeness. Standing there, my gaze shifting between these strange faces, I was regularly made conscious of myself as an indirect part of the work. In effect, the reflective properties of the installation physically served to blur the boundaries between myself and the work, including the portrait subjects. This encouraged me to question that relation itself: what was the connection between me in the gallery vis-à-vis the people represented on the wall – wherever they existed outside of the work? Contemplating this, I stood in front of *Spirit is a Bone*, devoting my full attention to the experience. After a while, it even felt as if the dark space in the work spread from the wall, engulfing me in it. In this sense, my experience at Hamburger Kunsthalle led to a moment of tangible identification with the group of people portrayed in *Spirit is a Bone*.

Upon arriving back home in Bergen, the holiday now over, the above experience still lingered in my mind. Impatient to make sense of it, I ordered the book, hoping that it might

give me even deeper insights. While awaiting its arrival, I set to investigate the project. In the following, I will give a general presentation of the work's production context, before going on to analyze the book.

Background

It turned out that my two strongest intuitions were correct: firstly, the portraits *were* inhuman in the sense that they were produced by an AFR system. Secondly, this system was specifically designed to capture people's likeness without their knowledge or consent (O'Hagan 2016). What Broomberg & Chanarin themselves refer to as *non-collaborative portraits* were in other words exactly as sinister as I had intuitively perceived them to be. Regarding the idea for the project, *Spirit is a Bone* turned out to have originated from a commission by the state-owned Russian media agency RIA Novosti, who wished for Broomberg & Chanarin to produce a series on labor in contemporary Russia. Somewhat skeptical towards the incentives behind this request, however, they replied with a counterproposal. They offered instead to do the commission on the condition that they could structure it as a contemporary Russian version of August Sander's *People of the 20th Century.* Additionally, the artists insisted that the subjects should be photographed using the facial recognition system FaceControl 3D by the Russian company Vocord (Mirlesse 2014: 39). In this way, what might originally have been conceived as a kind of propaganda project by the media company was turned into a form of artistic protest or political commentary.[30] To Broomberg and Chanarin's great surprise, Ria Novosti accepted their proposal, and the work began.

August Sander is known for his many portraits of people, which he photographed over a period of many decades from the 1910s until his death in 1964. His ambition was to document the German people, thinking that portraits of individuals would represent their ascribed class or social group as a whole. In the Preface to *Antlitz der Zeit*, Sander's first publication relating to this project, the art historian Alfred Döblin explains the artist's line of thought: from a distance, he argues, a person's individual traits disappear and the viewer will rather perceive the photographs as *en bloc* representations of his or her social group (Döblin 1983 [1929]). This idea reflects a belief in a correlation between outer appearance and "inner" qualities. Consequently, Sander categorized his portraits by denoting the sitters' profession or social

[30] It should be noted that I have not been successful in finding any empirical evidence for this claim. But as Russia is not known for its free press, however, I do not personally find the artists' skepticism in the matter controversial. In any case, Ria Novosti's original motivation is of no great consequence for the following discussion.

group rather than by name. When Broomberg & Chanarin were making their own version of his work, they therefore took to the streets of Moscow in order to find people that fit the same categories as those found in *People of the 20th Century*.[31] This included professions such as barber and paper manufacturer, but also more culturally relative categories, such as "elegant woman". The duo even managed to find a Dadaist whom they included. These people were then invited to a provisional studio in which a system on loan from Vocord was set up for the occasion. As the artists explain in an interview, the process from that point on was straightforward: "the subject merely had to walk into the room and the portrait was complete" (Mallonee 2016). This is very telling regarding the impersonal nature of the encounter between the portrayed people and the "photographer". In order to understand what happened during these few seconds, however, a short presentation of how the system operates is necessary.

According to the company itself, FaceControl 3D can be used for border control as well as crowd surveillance. It is especially suited in areas where large amounts of people are passing through, such as subway stations, airports or stadiums, and would typically be placed at entry points. Technically, FaceControl 3D works through four synchronized stereo cameras which record an area from different perspectives. From the image flows, the system is able to pick out and focus on individual people. Based on an analysis of a person's facial contours, it then proceeds to build a model of his or her face (VocordCompany 2016). Such a model can best be understood as a three-dimensional topography where the contours represent the individual's facial features (Fig. 21). A person's biometricized face is thus read almost like a map, and to the program it presents itself as no different from any other surface. Finally, the completed model is compared to images in the system's database to see if there are any matches. If so, it is reported to the operator in real time (Tadviser), who decides what to do with the information.

Regarding the meaning of the term *three-dimensional* used here, I adopt it from art historian Frank Popper, who in an early discussion of computer art distinguishes between two- and three-dimensional imagery. While both are presented on screens and are thus technically two-dimensional, the latter term refers to pictures based on algorithms for modelling three-dimensional effects (1993: 78). The images in *Spirit is a Bone* are three-dimensional in this sense. As the images are presented in the works on flat, immobile surfaces – as if photographic portraits – they must, however, have been subjected to a secondary process. First, I wondered if maybe Broomberg & Chanarin had done this themselves, but promotional videos published

[31] It is unclear exactly how Broomberg and Chanarin managed to find these contemporaries of Sander's sitters. While some of them are famous, such as Pussy Riot member Yekaterina Samusevich (revolutionary) and the poet Lev Rubinstein, the rest are ordinary, non-famous people.

by Vocord on YouTube make it clear that this remediation is a secondary function of FaceControl 3D: After building the topographical model, the system synthesizes stills from the image flow so that it becomes one coherent image, which it then "wraps" around the model to resemble a photographic portrait. The use-value of this step is unclear, but my guess is that it is designed for the convenience of the human operator. In this sense, they resemble what Harun Farocki would call *phantom images*: images made by computers for computers (2004).

The glitch-like qualities described earlier are thus the result of a many-step collage-like process in which a person's face has first been 1) filmed, then 2) abstracted from the image flow and 3) transformed into a three-dimensional model of nodal points, which is finally 4) covered with a photo-assemblage so as to look more recognizably like a face. Those models are still three-dimensional, however, so the images presented in *Spirit is a Bone* have further been 5) remediated by the artists into two-dimensional stills in order to be presented in book- and installation formats. In this sense, the final portraits can probably be understood as some kind of screenshots of the models as they were displayed on the system interface.

It is clear that the technical processes involved in transforming the images into the final work were far more complex than what was indicated in the above cited interview. It is also clear that FaceControl 3D differs on a technical level from the previously discussed algorithms. AFR systems can be divided into two general types at the level of their applications: those that use static images of the face and those that analyze dynamic images of faces from video (Gates 2011: 18). While ImageNet Roulette and *How do you see me?* were concerned with the former, *Spirit is a Bone*, on its side, takes issue with the latter. When curating the final work, then, Broomberg & Chanarin's work has seemingly consisted in choosing the particular angles of each computer model to present. The result of it all was the book *Spirit is a Bone*.

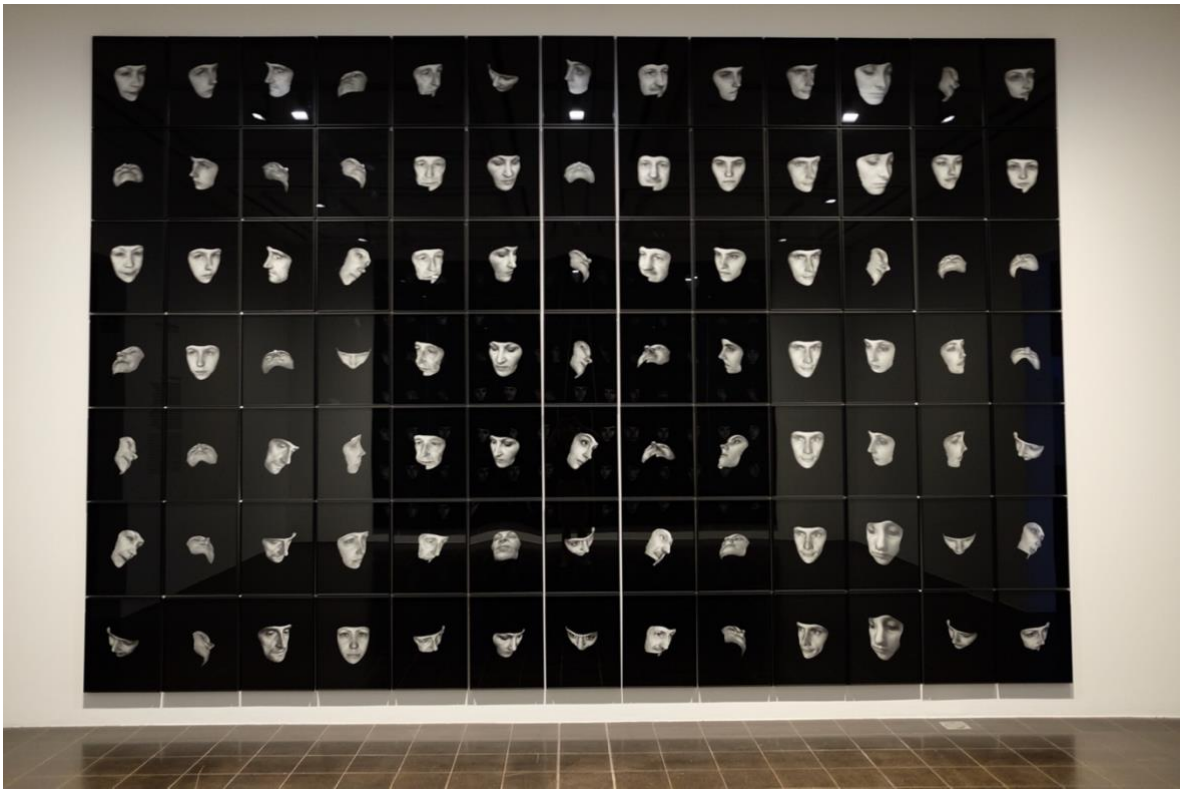


Figure 20: Spirit is a Bone *(2013) by Broomberg & Chanarin as presented in the exhibition* [CONTROL] NO CONTROL *at Hamburger Kunsthalle in 2018. Photo: Tuva Mossin.*

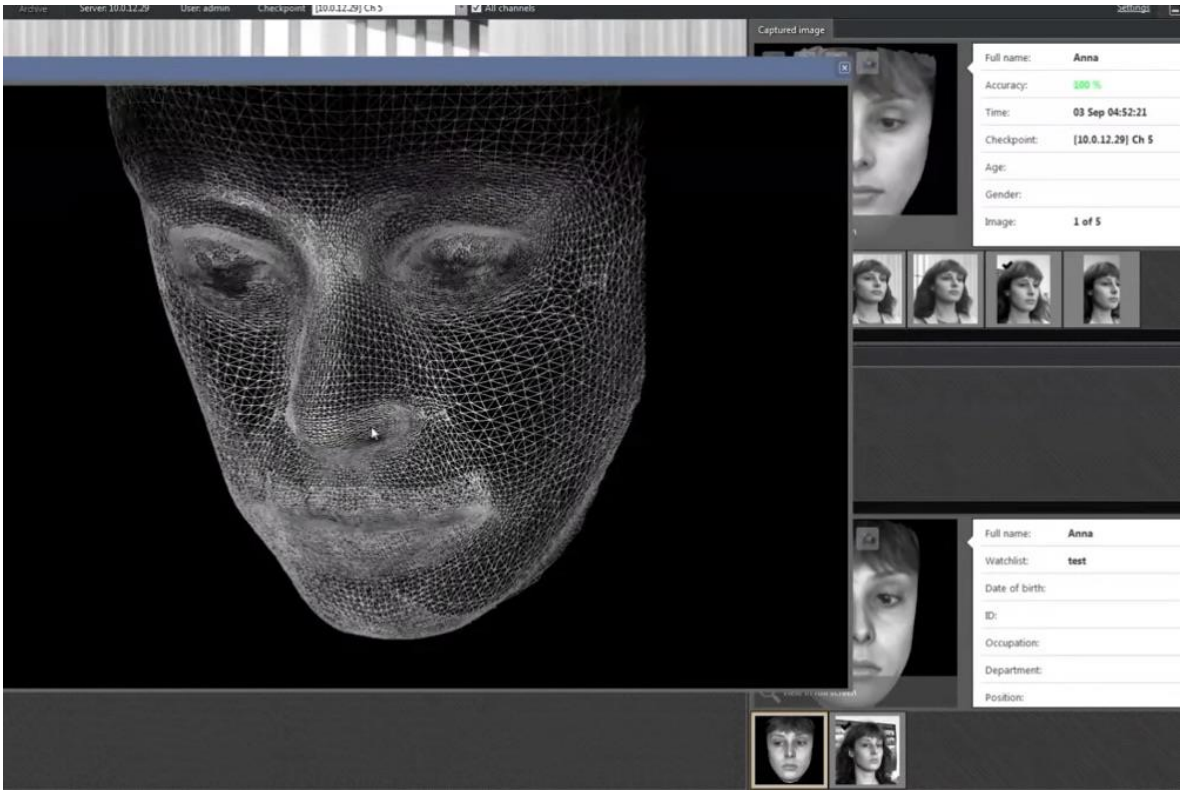


Figure 21*: Biometric model as it would appear to the FaceControl 3D user on the computer screen. Screenshot from audiovisual advertisement. URL: https://www.youtube.com/watch?v=FRq0qXOoBAc*

Second Encounter: *Spirit is a Bone* as Book

Finally, my order arrived by mail. The book was rather small and bound in white leather. The title was printed on the spine, while a woman's face in the well-known black and white, glitchy style was left to speak for itself on the front page. The back was also adorned with a print; this one of a face looking up so as to be seen from a frog's-eye view. When opening the book, I was presented with roughly 200 pages following the same layout: one mask-like face printed on the right side, while the left side had been left blank except for four lines of text printed in the bottom left corner. These were the categories written, respectively, in German, English, French and Russian: "Country Girl," "Young Farmer," "Confirmation Candidate," "Rural Bride," "Fighter or Revolutionary," "Widow" – these categories were all immediately recognizable to anyone acquainted with *People of the 20th Century*.

Even though the faces printed in the book were all of the same kind as the ones known from the installation, some noteworthy qualities nevertheless set the two versions apart. Besides, of course, the general differences between a book and a wall-hung installation, the most apparent amongst these was how the book's images were printed on a white paper background as opposed to black glass. Next, the paper's grainy quality made the faces, which were printed in a glossy finish, stand out from the background. This had the effect that when light hit the page, it danced around the photos' reflective edges while the paper stayed matte. Where the mask-like faces presented in Hamburger Kunsthalle had the appearance of being part of their dark surroundings, the material contrasts in the book had an almost opposite effect; rendering the faces doubly detached from their bodies and surroundings alike.

This trait, along with the temporal qualities of the book medium itself, invited a more intimate approach to each face: being able to look at only one page at a time, I was led to focus my whole attention to every face consecutively. In this, the book facilitated a closer connection to every single image than what I had experienced facing the installation – in both the physical and psychological sense. Where I had contemplated the portraits in the installation primarily on the basis of their expression as a group, the properties of the book led me to consider the subjects as individuals first, and only secondarily as a group. Next, the presence of individual titles also differed from the installation where the title *Spirit is a Bone* denoted the assemblage as a whole. The connection to Sander was therefore only evident in the book. The question still remains, however, what this appropriation of Sander's image universe adds to the work.

Figure 22: *August Sander*, The Woman of the Soil *(1912). Retrieved from MoMA.org on 28.05.2021. URL: https://www.moma.org/collection/works/193637.*

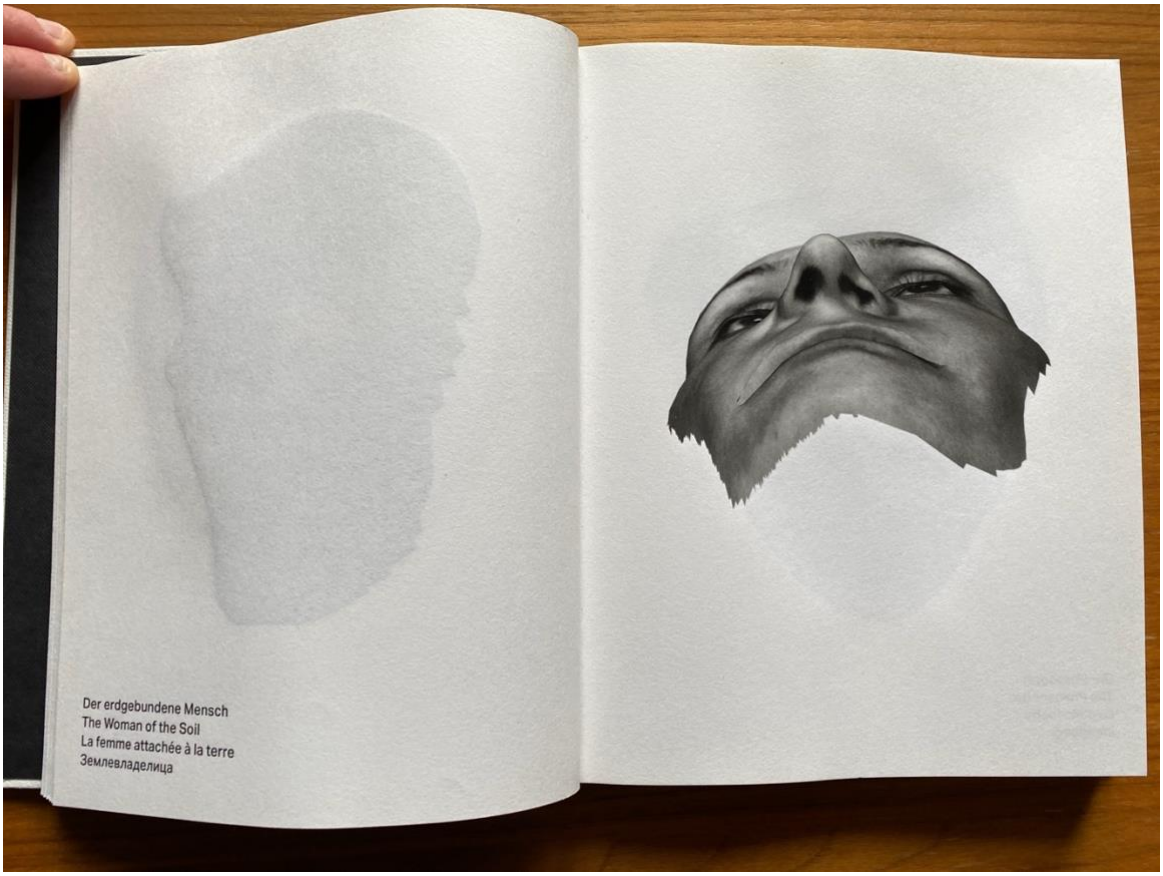


***Figure 23**: Broomberg & Chanarin,* The Woman of the Soil, *in* Spirit is a Bone *(2015). Photo: Tuva Mossin.*

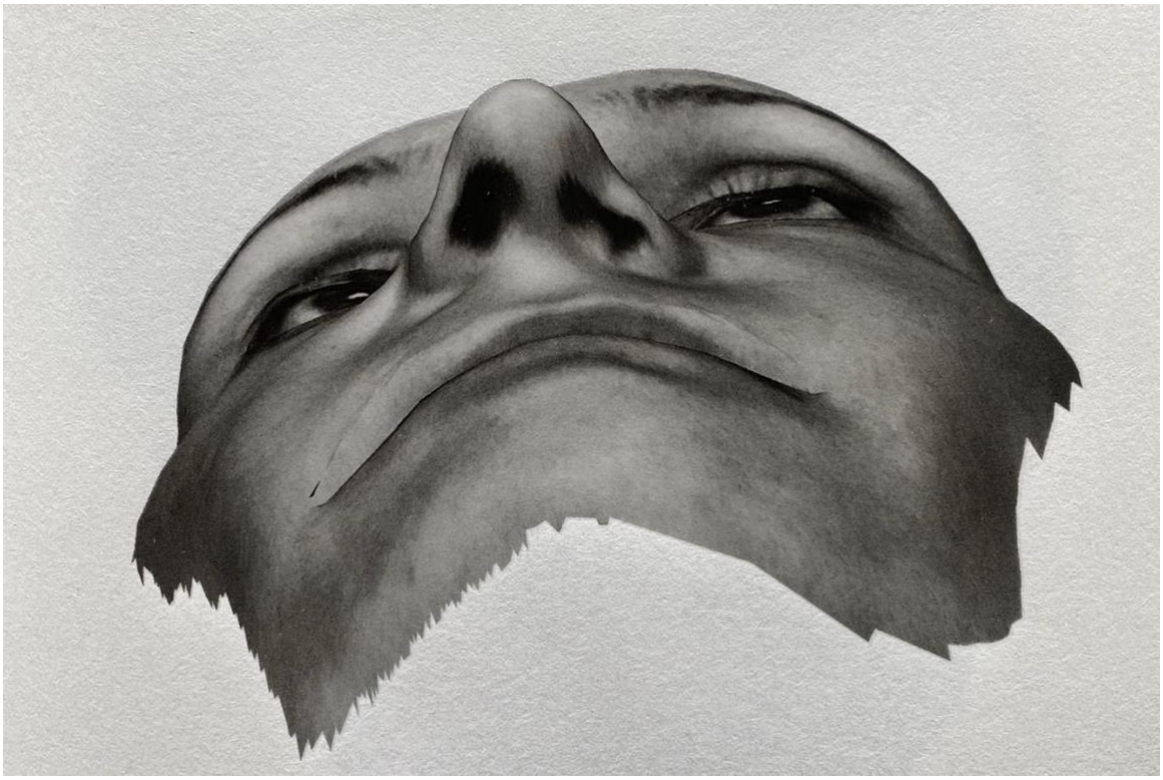


***Figure 24:** Broomberg & Chanarin,* The Woman of the Soil *(detail), in* Spirit is a Bone *(2015). Photo: Tuva Mossin.*

Apart from the organization according to Sander's categories, the aesthetic experience of looking at *Spirit is a Bone* was very different from that of *People of the 20th Century*. While Sander usually portrayed his subjects frontally, either sitting or standing in surroundings meant to reflect their life situation, the portraits in *Spirit is a Bone* revealed virtually nothing indicative of the subjects' circumstances. Take Sander's portrait *The Woman of the Soil* as an example (Fig. 22). This photo, taken in 1912, shows an elderly woman dressed in black. With her hands folded around a book, she is sitting in a robust armchair, wearing a loosely tied headscarf. The composition fills about four fifths of the entire image, while the top fifth has been left as dark as the woman's attire. Even though she is only portrayed from the knees up, the parts included are able to express much about the sitter. Her wrinkled face and hands – the only parts of her body not covered by clothes – tell not only of old age, but also of a lifetime of physical labor. Her black clothes and covered head indicate piety – leading me to assume that the well-used book in her hands must be a copy of the Bible. It is her face, however, which really brings the composition together for me: staring back at the viewer with a stern look, a slight smile can also be found on the woman's lips. Somehow, I read from this face that even though her life has made the woman sturdy, she is nevertheless friendly and contempt – at peace with her circumstances as well as with her God.

Whether these assumptions are correct or not, the point is that this experience is radically different from the contemporary Russian counterpart in *Spirit is a Bone* (Fig. 23-24). This portrait shows one of the facial models tilted at an angle as if the woman is looking upwards. From such a dramatically foreshortened perspective, it is practically impossible to read the expression on the face – if there even was one to begin with. Looking up at the face from below, the bottom of the nose becomes the central focus, whose uneven nostrils draw attention to the image's twisted nature. Furthermore, the edges of the face have been especially crudely cut in this image, leaving the bottom edges jagged like a ripped-up fabric. Had it not been for the title, it would have been all but impossible to guess anything about the depicted person.

Needless to say, the three works presented above, *People of the 20th Century* and the two versions of *Spirit is a Bone*, were grounds for quite different experiences. Standing in front of the installation, I had an experience of being enveloped in the same digital space as where the faces were floating, which in turn aroused a curious feeling of being monitored myself. Apart from the same glitchy qualities of the faces themselves, the experience of browsing the book did not engender any associations to the Internet. It did, however, give rise to an uncomfortable feeling along the lines of reading someone's diary – as if by looking at these people's likenesses, I was crossing some moral line, invading their privacy. *People of the 20th*

Century, on its side, is a series of portraits equally expressive as the one described above. The remaining question, then, is what do Broomberg & Chanarin gain by inviting a comparison through appropriating Sander's categories? And why did they choose this photographer in particular?

The roughly one hundred portraits presented in the book version of *Spirit is a Bone* are followed by a literary conversation between the artist duo and Eyal Weizman, the architect known for leading the multidisciplinary research group Forensic Architecture. The text is titled *The Bone Cannot Lie*, and discusses *Spirit is a Bone* in light of the twin pseudo-sciences phrenology and physiognomy. Both of which had the human skull as their central object of study. As goes from the above presentation, this is also true of FaceControl 3D in that the digital representation of physical features are essentially maps of the head stripped down to the bone. As Weizman describes it, "the shift to the bone [in phrenology] signified a certain unveiling, stripping down to the essence, in the double meaning of the term" (Broomberg & Chanarin 2015: 218). Facial recognition technology based on digital biometrics also seeks to "peer past" the layer of skin – with all its confusing make-up and accessories – to determine the underlying shape, as this is seen as a more accurate than other available methods.[32]

From this conversation, it becomes clear that Broomberg and Chanarin's intention when making *Spirit is a Bone* was to invite reflection around the ideological parallels between these historical ways of thinking and the ideas behind the use and production of automated facial recognition technology today. *The Bone Cannot Lie* also draws parallels between the contexts in which the historic and contemporary technologies are implicated: the identification of suspect individuals. The three also hint towards the fact that, in a broad sense, the spread of physiognomy is closely connected to later racist theories such as Social Darwinism, anthropometry, as well as the "racial science" used to legitimize genocide by German National Socialists (e.g. Morris-Reich 2016). That these parallels are explicitly emphasized in *Spirit is a Bone* therefore deserves serious consideration. I will thus take a detour to give a presentation of physiognomy and other positivist methods of identification before getting back to Sander and how his oeuvre relates to the overarching discussion. As will be made clear, the history of photography was from the medium's very invention closely related to physiognomy, which in

[32] This is indeed how Vocord has advertised the advantages of its own facial recognition as opposed to competing two-dimensional systems whose accuracy they criticize for plummeting when the subject changes pose, wears glasses or holds things like a cigarette in front of their face. As opposed to these, their own system "won't be fooled by hats or scarves. Attempts to disguise skin color may fool many 2D systems, but can't trip up FaceControl 3D" (VocordCompany 2016: 01:56).

turn became closely related to criminal identification. The implication of a famous photographer in this story is therefore not at all random.

Photography and Physiognomy

While physiognomy has roots dating back to antiquity, the discipline first gained widespread interest in the eighteenth century, reaching its peak in the late 1770's when Johann Kaspar Lavater's (1741-1801) published *Physiognomische Fragmente zur Beförderung der Menschenkenntnis und Menschenliebe*. The Swiss' motivation was to create a scientific framework with universal appeal through which to investigate the relationship between a person's appearance and character (Mortensen 2012: 39). In the studies Lavater presented here, he examined people's cranium, profile, skin color and other facial features, and tried to connect these to corresponding personality traits. In this sense, physiognomy was treated as a system of signs with fixed decipherable meanings (Mortensen 2012: 41).

While the "physiognomic era," as some call it, was well past its glory days by the time photography was invented, media scholar Mette Mortensen has argued that it gained a new flowering as a direct consequence of this new medium. A central problem that the discipline had faced before was the dependency on drawing, which rendered the documentation of physical traits very slow and tiresome. The photographic portrait seemed to solve this particular issue, but this was not the only reason that the medium was incorporated into a physiognomic framework. As Mortensen argues, this was probably just as much a result of the fact that physiognomy, more than serving a methodological need, offered a vocabulary for discussing a sensual experience of the daguerreotype as an authentic and true image: on the one hand, portrait photography granted the physiognomists new objects of study, while on the other, the pseudoscience offered an explanatory model for this new representational genre which at that point had no interpretive framework (Mortensen 2012: 33-34).

It might also be, Mortensen argues, that the photographical portrait's aura of objectivity served to assure these pseudo-scientists – which had by then been widely discredited – of their own scientificity (2012: 36) In this sense, photography's openness to interpretation paradoxically rendered it a fruitful surface upon which to project one's own prejudices. Implicated in a framework that treated the photographic portrait as a signifier of the depicted person's true self, it is not surprising that when the growing European bureaucracies saw the need for more trustworthy methods of criminal identification, this medium appeared especially well suited.

Photography was applied to systematic cataloging of arrestants already by the 1860s, and the Frenchman Alphonse Bertillon (1853-1914) is usually credited as the originator of the standardized "criminal image." The so-called bertillonage was characterized by a radical lack of individual adaptation. Much like the contemporary mugshot, every arrestant was to be depicted twice: frontally and in profile (Fig. 29). In addition, a range of anthropometric measurements were made so as to leave very little room for individual interpretation. In this sense, the bertillonage was an unconscious break with physiognomy in its mistrust of the face to speak for itself. It also constituted a break with criminal depiction by commercial portraitists, whom Bertillon criticized for being too focused on fashions and retouching in line with their usual costumers' needs (Mortensen 2012: 102). On the contrary, he sought methods of exposing characteristic traits as much as possible, and the answer he came up with was to rid the picture of background details, deny the wearing of head accessories, and to insist on neutral facial expressions and gestures. By 1885, these methods had been implemented across France. American police followed suit in 1888, and several other European countries soon followed. Now, law enforcements agencies could easily compare records across national borders.

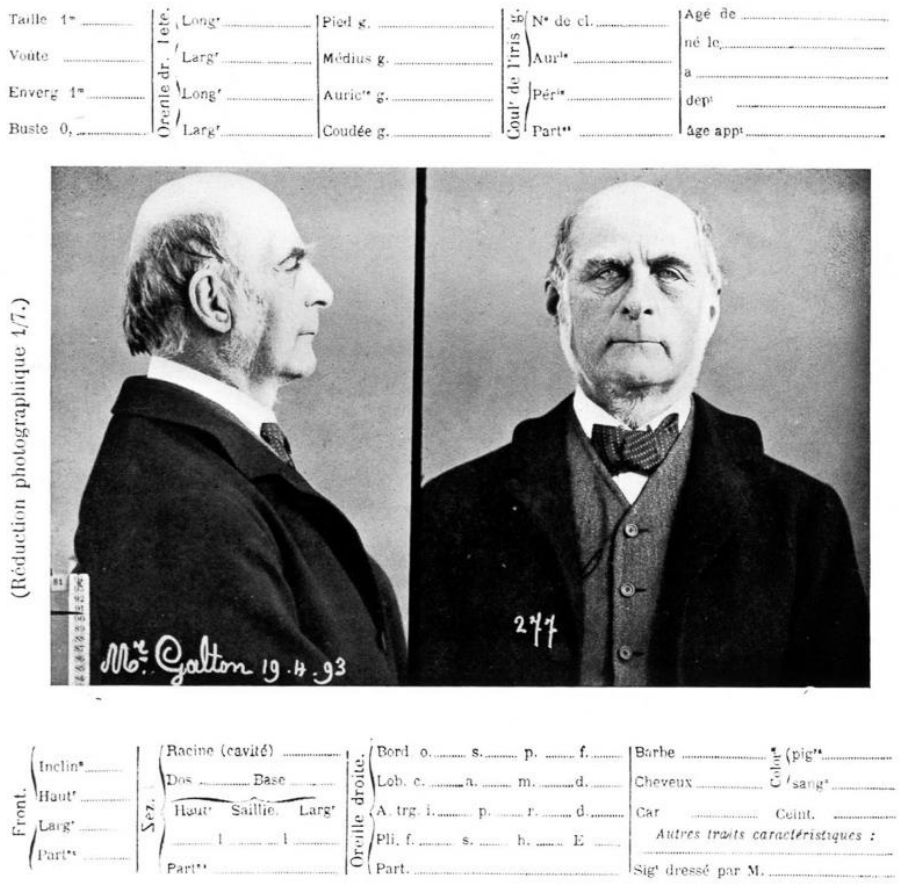


Figure 25*: A typical bertillonage depicting Bertillon's colleague, Francis Galton. Retrieved on 28.05.2021 from educalingo.com. URL: https://educalingo.com/en/dic-de/bertillonage.*

Bertillon was not alone in his use of the photograph as empirical material, as he was deeply influenced by the methods of anthropologists in their study of non-Western peoples and races. In the essay *Tracing Photography*, Elizabeth Edwards discusses how photography has historically been used to establish anthropological facts. As she tells it, photographs early on became privileged sites for communicating a feeling of immersion in foreign cultures to the people "back home" who could not themselves take part in the fieldwork in foreign lands. This resulted in the development of elaborate conventions for visual documentation. One example is the idea that the photograph had to be "spontaneous" so as to eliminate posing: if the subject stopped to pose, then the scientist had supposedly crossed the border of acceptable participation and undermined disciplinary truth. To situate a whole situation itself, however, was paradoxically deemed to be within acceptable parameters (Edwards 2011: 162-64).

This skepticism towards posing can be understood as analogous to how Bertillon sought to rid his criminal portraits of gestures and facial expression to appear more neutral. But the photographic apparatus was not the only tool used by scientists in second half of the 19th century to document differences within and between ethnic groups: this was also the age of craniometry, but even more importantly, it was the age of Darwinism. In 1859, twenty years after the launch of the first photographic techniques, Charles Darwin published *On the Origin of Species*. Evolutionary theory swept the continent, influencing most disciplines in one way or another, but especially those who had human beings as their object of study. Simultaneously, a strong belief in the potentials of mathematics was in vogue. The paleontologist and evolutionary biologist, Stephen Gould, describes the general mood of this period captivatingly:

> The second half of the nineteenth century was not only the era of evolution in anthropology. Another trend, equally irresistible, swept through the human sciences – the allure of numbers, the faith that rigorous measurement could guarantee irrefutable precision, and might mark the transition between subjective speculation and a true science as worthy as Newtonian physics. Evolution and quantification formed an unholy alliance; in a sense, their union forged the first powerful theory of "scientific" racism – if we define "science" as many do who misunderstand it most profoundly: as any claim apparently backed by copious numbers (…) By the end of Darwin's century, standardized procedures and a developing body of statistical knowledge had generated a deluge of more truthworthy [sic] numerical data (Gould 1997: 105-6).

It is clear that the drive to submit individuals to rigorous measurement was not exclusive to Bertillon. Nor was his belief in the truth-potential of developing standardized procedures. I also find it worth noting how Gould mentions that scientists from the humanities held the natural sciences in such high esteem, as this is especially illustrative of the respect towards mathematics

as grounds for analysis. When these two trends intersected, it became breeding grounds for a series of problematic and oftentimes absurd practices. Of all those that could be cited, I find craniometry to be the most relevant to the discussion in this chapter.

The Paul Broca School, with the French physician, anatomist and anthropologist (1824-1880) as its front figure, was the predominant branch of craniometry. As the name implies, craniometrists focused on the skull and brain, which they measured meticulously in order to develop a system where racial and social groups' average measurements were connected to their assumed psychological, physical and intellectual abilities. As Broca himself put it in a debate, they were devoted "almost exclusively to the study of skulls," from whose data they "hoped to find some information relevant to the intellectual value of the various human races" (Broca 1861: 141). Not only did the craniometrists assume that they by numeric measurements of the brain's different areas would find out something interesting about intellectual properties per se. Inspired by evolutionary theory, they simply assumed that the parts that are bigger in the human brain than in other animals must surely be the center for higher intellectual properties – and thus, of course, at their most developed in white males: "in general, the brain is larger in mature adults than in the elderly, in men than in women, in eminent men than in men of mediocre talent, in superior races than in inferior races (…)" (Broca 1861: 304). From the outset, then, there was inherent misogyny, ageism and racism within the discipline, but the ranking of races was nevertheless the discipline's most essential goal.

Craniometry was belief in numbers taken to an extreme, and it was not only the musings of some flamboyant Victorian scientists in exclusive journals. Conclusions from craniometric studies flooded the popular press and thus had a wide impact on society at large. In at least one case, an American journalist concluded that there was scientific proof that people of color should not be given voting rights because of their supposedly smaller genus – a find which was soon after disproven (Gould 1997: 114). But despite the political consequences, the leading figures of craniometry were not conscious political ideologues. On the contrary, they regarded themselves as servants of their numbers, apostles of objectivity. Even so, as Gould puts it, "they confirmed all the common prejudices of comfortable white males – that blacks, women, and poor people occupy their subordinate roles by the harsh dictates of nature" (Gould 1997: 106).

As with the physiognomists, a definitive pattern is discernable in that the gap between empirical data and conclusions is systematically reversed: the conclusions come first, and the data is collected selectively so as to best confirm them (Gould 1997: 117). The creative interpretation of numbers is especially apparent in an example where it became clear to Broca that some Asians, or "yellow people," actually possessed larger skull capacities than Europeans.

Instead of admitting that either the concept of ranking races on a linear, hierarchic scale was futile from the beginning, or that Europeans were not on top, he exercised some impressive interpretive acrobatics to be able to stick to his preconceived ideas (Gould 1997: 119).[33] While marching under the banner of empiricism (Broca's disciple, Paul Topinard even took as his motto "I abhor systems, especially a priori systems") even their own data could not overthrow their initial assumptions. It is also evident, on a systemic level, that the questions that are asked in the first place are without exception of the kind that serve to affirm the existing social order placing white males of a certain social standing on top. This pattern is perhaps the most apparent within criminal anthropology, exemplified by the work of Cesare Lombroso (1835-1909) who founded the Italian School of Positivist Criminology.

Lombroso proposed that crime is a characteristic of human nature to which some are more disposed than others. The "criminal type" could be recognized by atavistic physical and psychological characteristics which he deemed "apelike" (Fig. 30). While "normal" people could also perform criminal *behavior*, only some were hereditarily disposed to it. In his research, Lombroso sorted criminals into categories such as "shoplifters," "purse snatchers" and "swindlers," and claimed that the individuals in each category shared certain physical traits, or *stigmata*, as he called them.[34] Not surprisingly, these early criminologists also held that "gypsies," as well as "savages" from other continents, were less developed than Europeans from an evolutionary perspective, and claimed to demonstrate this by pointing to their criminal-like behavior and general shamelessness (Gould 1997: 142-75).

In reality, these tendencies – which had little empirical support – were simply observations of cultural difference from a standpoint of moral objectivism: whatever appeared unmoral to the European observer was taken to be a sign of unmoral leniencies amongst other peoples. While Lavater's methods were similar to craniology (largely consisting in measuring diseased bodies, but according to much less standardized schemes than the Broca school), it was his theory of *atavism* that was most controversial during the period. The heated debate about whether there was a biological factor in criminality was very important around the turn of the twentieth century, and although it was highly criticized and later debunked, it was nevertheless widely influential. It should also not go without saying how perfectly well suited the idea of hereditary criminality is for conservative political argument: if the situation of the poor, "stupid," and degenerate is the result of their own unfortunate dispositions rather than

[33] I discuss this example in more detail in Chapter 4.
[34] Examples of such were darker skin, epilepsy, and even a knack for using onomatopoeia in speech.

socio-economic circumstances, then the status quo – even if unfair – simply reflects nature. In other words, this was naturalization of politics, which is any ideology's ultimate trick.

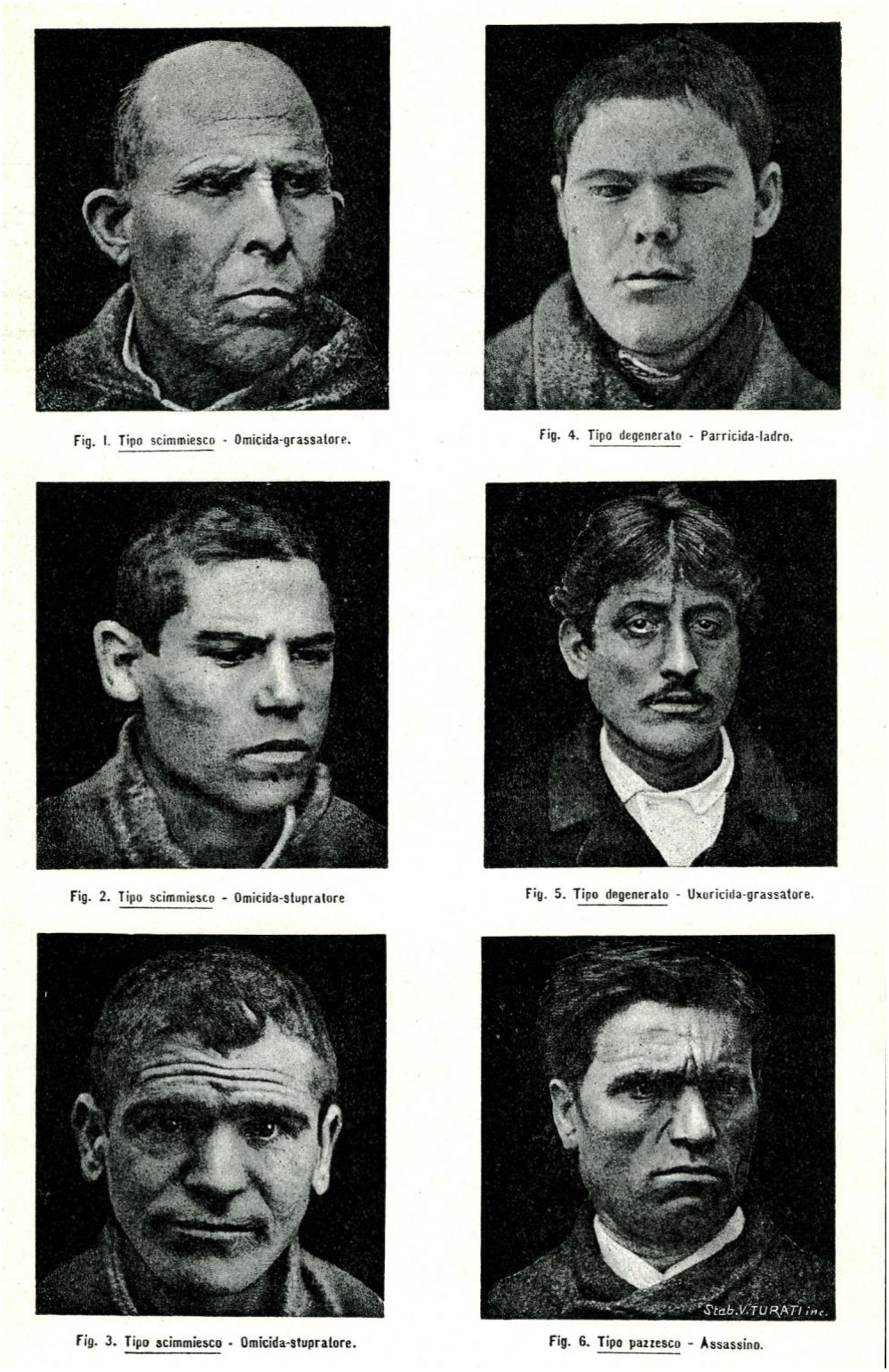


***Figure 26**: Examples of criminal types presented by Lombroso in his main work* L'uomo delinquente *(1876). Retrieved from Wikimedia commons on 28.05.2021. URL: https://commons.wikimedia.org/wiki/File:L%27uomo_delinquente,_1897_%22Tipi_di_Omicida...%22_(415782 7276).jpg*

So was the context in which the establishment of criminal photography took place. In Europe, the police started to photograph arrestants systematically in the 1860s. This was partly a response to the rapidly growing urban populations and industrialization, which led to a more mobile citizenry. Previously, local police officers had likely been well acquainted with the people within their ascribed jurisdiction, but the situation was very different in densely populated urban areas with foreign people relocating from other places or merely traveling through. This led governments to seek ways through which to control what was perceived as growing security threats in a modern society full of aliens. Comprehensive archives were thus created in which states collected all kinds of data about their citizens. Important to this chapter's discussion is the two-fold nature of these institutional developments: on the one hand, the "right" identity gave you access to certain rights such as freedom of movement across borders. On the other hand, a criminal record or inability to verifiably document yourself could be used against you. Indeed, the national identity archives were widely used to prosecute criminals (Mortensen 2012: 66).

Identity portraits, then, were from the beginning ingrained in the institutional set-up of the newly conceived European nation states. Simultaneously, the growing industrial bourgeoise were the portrait photographer's main costumers. As professional portrait photographers were often the same ones who also documented arrestants, the early staging of the "criminal body" built on the visual language of bourgeois visual representation. Mortensen describes this balancing act between representing bourgeois identity positively while also differentiating those who deviated from those ideals in the following way:

> (…) a new bourgeoisie emerged in Europe and USA, who saw the photograph as a nearly perfect showcase for its own values and moral dogma. But the medium of photography did not only help the bourgeoisie in visually formulating its identity. The criminal photograph also serves as an example of how it was used to put a face on those people who for either social, political, moral, medical or ethical reasons deviated from the bourgeois norm (2012: 65).

When, as mentioned in Chapter 1, Crawford and Paglen wrote that the system they critiqued was similar to these nineteenth century disciplines in essentially measuring "deviance from bourgeois ideals," this is what they meant: a systemic pathologization of what was seen as criminal behavior across a wide specter of institutions serving the state and implicitly maintaining the existing social order.

In *The Burden of Representation,* John Tagg discusses the photograph as one tool in a larger structure of domination which by the turn of the twentieth century constituted an

institution of knowledge in its own right. The photographs taken in institutional contexts – the carceral system as well as others – must therefore be treated as records of the power yielded by the state: "each device is the trace of a wordless power, replicated in countless images, whenever the photographer prepares an exposure in police cell, prison, mission house, hospital, asylum, or school" (1988: 64). This complex constellation of institutions sought to construct docile subjects through various disciplinary techniques operating "unremarked in the smallest duties and gestures of everyday life" (Tagg 1988: 62).[35]

Besides being used as a subjectivizing tool, there is another shared aspect between phrenology and AFR: they are both constructed to serve a *prohibitory* function. This is commented by Weizman in *The Bone Cannot Lie*: "we find ourselves in a reality analogous to phrenological principle of prediction looking at various patterns and form to see into the future" (Broomberg & Chanarin 2015: 228). Taking into consideration the very specific contexts in which surveillance systems like these are being employed, it is obvious that FaceControl 3D and its sibling-systems are also on the lookout for criminal types. As is also pointed out in the text, it is interesting how the archive, time and again, appears central in deciding who is targeted by these systems.

I have already described the methods of Lombroso and his colleagues. Therefore, the parallels should be clear as day when researchers at Harrisburg University in Philadelphia declared in 2020 that they had developed a system that could "predict if someone is a criminal, based solely on a picture of their face" – a capacity they advertised for use in law enforcement to predict crime (BBC 2020). This was in the USA, where it is a well-established fact that people of color (POC) constitute a disproportionately large number of inmates due to racial profiling by police (e.g. Balko 2020). With Chapter 1 in mind, it is therefore easy to imagine that if an algorithm is trained in an archive of mugshots, then already disadvantaged groups will be unjustly profiled by it. But even if there existed no such disproportion, one is nevertheless left with the fact that *any* predictive system is trying to foresee the future based on the past, as statistics are always retrospective in the sense that models are based on historic events.

Statistical prediction is thus essentially deterministic in its repetition of history. However, the opposite – a truly psychic system – would be no less unproblematic. This idea is the basis of Stephen Spielberg's movie *Minority Report* (2002) based on a novel by Philip K. Dick. Here, law enforcement has developed a system which foresees crime already before it

[35] As Tagg also notes, such invisible and diffuse forms of subjectivation were the same as the microphysics of power described at length by Michel Foucault in *Discipline and Punish* (1975).

takes place, and their job therefore consists in hunting down forthcoming criminals and convict them before they have the opportunity to commit the crime. The story takes a turn when John Anderton, the main character played by Tom Cruise, becomes a victim of the "PreCrime" system himself, prosecuted for a murder he has yet to commit. The story from there is a series of events reflecting on the dangers of such presumptive judgement. The relevance to criminal anthropology is not so far-fetched, as such measures were indeed proposed by one of Lombroso's supporters in 1897: prescreening and isolation of people bearing stigmata *before* they had committed any offences (Gould 1997: 166).

In contrast to this science-fiction universe where Tom Cruise's character belongs to the dominant social class, however, the injustices of such a system in reality would most likely be even more worrisome. As the discussion so far has demonstrated, methods of predictive policing almost inevitably serve to uphold the status quo of existing socio-economic hierarchies and favor the ruling classes. This tendency is in line with the research of gender studies scholar Shoshana Magnet, which shows that ethnic minorities and subaltern groups such as transgender communities are disproportionally profiled in biometric systems: "clearly biometrics allow the state to both implicitly and explicitly engage in racial profiling while using the rhetoric of technological neutrality and mechanical objectivity to obscure this fact" (2018: 118).

August Sander Revisited

Sander's work is relevant to the above discussion in several ways. Through his lifelong photographic project, he aspired to create an exhaustive catalogue of the German people. Thinking that by depicting people in a certain way they would present as icons or archetypes of their ascribed group, the overarching goal could be achieved by depicting people of all relevant groups. In this sense, his belief in photography's ability to express "human nature" was a continuation of physiognomic thought. Like those seeking to capture the visual appearance of criminal types, Sander apparently also held a strong belief in the value of categorizing people meaningfully into groups. This being said, he might not have believed in a completely one-to-one relationship between someone's facial anatomy and their "character." Nevertheless, he developed his project within what photographer and theorist Alan Sekula has described as the "legitimizing aura of physiological terminology," meaning that Sander wrote and talked about his own work in terms reflecting the terminology of said discipline (1981: 18).

In light of this, it might be tempting to start searching for other reactionary tendencies within Sander's work, but the assumptions regarding the relationship between physical appearance and "inner" qualities was not at all limited to the political far right. On the contrary,

Sekula argues, both phrenology and physiognomy were considered legitimate academic disciplines across the political spectrum and thus cannot be attributed to any particular political ideology. A more fruitful question to ask is thus rather what purposes the theories were meant to serve. As Sekula has argued elsewhere, photography became a preferred tool for positivists for two reasons. Not only because it was seen as objective, but also because photography, as he puts it, fulfilled "the Enlightenment dream of a universal language: the universal mimetic language of the camera yielded up a higher, more cerebral truth, a truth that could be uttered in the universal abstract language of mathematics" (1986: 17). While Sander was supposedly skeptical towards the biological determinism implicit in physiognomy, he was all the while a believer in what Sekula refers to as a "universal pedagogic" believed by the liberal left to have the potential to lead to a pan-human understanding of a "language of the body" (1981: 16). Not to say that this belief was without problems of its own, because it also seems to build on a technological positivism related to physiognomy.

Sander was, like most positivists, indifferent towards the epistemological differences between peoples and cultures, which were assumed to lie at the surface level only (Sekula 1981: 18). In this, he shared with the National Socialists the same positivist ideas that they used to legitimize their own fascist reign over other peoples (Sekula 1981: 19). It should be stressed, however, that Sander's world view in many ways also directly contradicted Nazi beliefs, as is demonstrated by his inclusion of Jews, revolutionaries and differently abled people in his visualization of the German people. The fact that Nazi officials actively prohibited and even destroyed much of Sander's production precisely for this reason in 1934 serves to demonstrate this (Clarke 1992: 74). But the main point still remains true, which is that Sander's project gave voice to the same institutional authority which legitimized physiognomy and the later "racial sciences."

In this respect, it is interesting how, when put next to *Spirit is a Bone*, his portraits do not appear as the purely empirical collection I have just indicated. As Sekula points out, he proceeded with his project without reducing the "aesthetic coherence and semantic ability of the traditional portrait form" (1981: 18). This means that, as opposed to how Lombroso depicted people in his research on criminal types, Sander never gave up the expressive aesthetic tools associated with portrait photography. *Spirit is a Bone* serves to highlight this fact about Sander, as the above comparison of *The Woman of the Soil* demonstrates. This point finds support in an essay by Graham Clarke, where he discussed exactly the two-fold nature of Sander's photographs. While on the one hand, the portraits are situated within a fixed social code and

closed realist aesthetic, they are also strikingly personal: "while he seeks types, he celebrates individuals" (Clarke 1992: 82).

Enter Biometrics

It is clear that all of the photographs presented in this chapter are in some way or other related to identity, either of an individual or of groups. In this sense, they can be understood as ID photographs; a genre existing in various forms, whereof passport photography and so-called mugshots are the most well-known variants today. I would argue that the images seen in *Spirit is a Bone* are most closely related to the latter – both practically and aesthetically. The mugshot is usually a double photograph depicting an arrestant frontally and in profile and is used within law enforcement to document suspects in criminal cases. The images created by Vocord 3D, on their side, have merged the profile and frontal view into one model. They are also practically related in that such photographs are not made of everyone, but only of those who for some reason or other have come to be deemed suspicious or as a threat to society.

While techniques such as fingerprinting and DNA with time came to complement photography in the process of criminal identification, a method for positively determining someone's identity completely free of ambiguity has never been achieved. This issue is emphasized in an article by the historians Carlo Ginzburg and Anna Davin:

> But there was a still more serious defect in Bertillon's anthropometric system, the fact that it was purely negative. It permitted the elimination of those whose details on examination did not match up, but it could not prove that two sets of identical details both referred to the same person. The elusive quality of individuality could not be shut out – chased out through the door by quantification, it came back through the window. So Bertillon proposed combining the anthropometric method with what he called a word-portrait, that is, a verbal description analyzing particular details (nose, eyes, ears and so on) which altogether was supposed to reconstitute the image of the whole person, and so to allow identification (1980: 25-26)

This is a good example of the same kind of desperate spiral which was evident in the discussion of ImageNet in Chapter 1. When Bertillon faced challenges regarding his identification methods, instead of questioning the practice itself, he assumed that the answer to his troubles must lie in the accumulation of *even more data*. As I will discuss in more detail in Chapter 4, the same tautology lies at the heart of modern identification techniques based on digital biometric systems. In this sense, the AFR systems discussed in this thesis can be understood as the latest developments in the more than two-centuries-long strive towards positive personal identification. It is evident that such a system has never existed.

As it is clear that the connotations inherent to the genre of ID photography has its flaws, I therefore propose that a fruitful reading of *Spirit is a Bone* would be to consider the portraits as *data doubles* instead. This term is used by the sociologists Kevin Haggerty and Richard Ericson to denote the transformation of the body into virtual information, or "data flows," through the various processes to which people are exposed by public and corporate institutions. Even though the digital doubles point back to specific individuals, they nevertheless fall outside the framework of traditional representation in their focus on use-value rather than mimetic resemblance. This is because this more effectively serves the particular institutional interests for which such models are built (e.g. state control or targeted advertising) (Haggerty & Ericson 2000). Such a reading can function to emphasize the changed nature of visibility brought about by global digitalization: while visibility has historically been considered a social good – a means of having one's voice heard – it is something to be avoided in the digital realm.

Art historian Melissa Gronlund comments in a discussion of digital art on how identity politics has for a long time been concerned with visibility as a means of political empowerment. She argues that this has changed with the invention of the Internet: online, visibility comes with certain risks, like rendering already marginalized groups increasingly susceptible to surveillance and datamining (Gronlund 2017: 7). While this is to some extent true, the history of the portrait renders this distinction less clear-cut than how Gronlund presents it. One aspect of the bureaucratization that came with the industrial revolution was that the genre of the portrait – which had previously been reserved for those who could afford to commission a painter or sculptor – was democratized with photography. Everyone's likeness was now to be documented and logged. This had the consequence of reversing the status of the portrait form itself: where it had previously been a privilege to be pictured, it was now "a burden of a new class of the surveilled" (Tagg 1988: 59). This dark side of technological development is therefore not new with the internet but is rather being replayed – even amplified – with the spread of digital biometric identificatory techniques.

Hegel: "The Spirit is a Bone"

The wide popularity of Lavater's physiognomy should not be taken to mean that his thoughts stood unchallenged by his contemporaries. Immanuel Kant, for one, was skeptical towards the idea that a person's physical attributes could determine their actions because this would imply a rigid determinism, which would contradict personal responsibility and moral. But it was the physicist Georg Christoph Lichtenberg (1742-1799) who stood for the most severe criticism of physiognomy. This is of importance because it brings me to the matter of the art project's title,

Spirit is a Bone, which I have not yet discussed. The phrase is taken from Hegel's *The Phenomenology of Spirit* (1807); the philosopher's most influential work where he describes the dialectical progress of the spirit from its primitive origin until it becomes "pure knowledge" through a series of stages. The phrase comes from the chapter discussing the stage where the Spirit reaches *reason* (chapter C(AA), Vb).[36]

Leaning heavily on Lichtenberg, Hegel goes to lengths to describe how the discipline of physiognomy (which he with great irony refers to as a "science"), is based on a tautology where man seeks answers in nature for ideas he has himself come up with. Whatever "laws" such a science finds, Hegel argues, can therefore be nothing more than an "empty opinion" (2009 [1807]: 179). This is because man's true being cannot be found in *matter* but must be sought in *action*, which is why is it a futile endeavor to seek it in the skull:

> The skull-bone is not an organ of activity, nor even a speaking movement; we neither commit theft, murder, etc. with the skull-bone, nor does it in the least betray such deeds by a change of countenance, so that the skull-bone would become a speaking gesture. Nor has this *being* the value even of a *sign* (2018 [1807]: 134).

When Hegel finally formulates the phrase "the spirit is a bone," then, it can be understood as a condescending way of summarizing this confusion of lifeless matter with "true" spirit. The phrase is only true for those people who have not yet progressed from that particular stage of reason, but not for those with a more advanced form of understanding, capable of contemplating things *dialectically*. It is not surprising then, that Hegel's fiercest opponents were the positivists who modeled their thinking on the mathematical natural sciences, and whose ideas his work can be read as a direct criticism of (Østerberg 2009: 12). So how, then, does this choice of title affect a reading of Broomberg & Chanarin's project?

Intuitively, I take the title to signify a kind of alliance by the artists with Hegel against positivism, of which modern biometrics is definitely a strand. On a deeper, more speculative level, I would like to argue that the title can also be understood as an ironic nudge towards his philosophy. As Hegel's idea of the spirit was conceived as striving towards its highest level of pure knowledge, the fact that the line of thinking he criticized still persists opens up the question of whether he was right. If mankind has apparently not progressed very much in the more than two centuries since the publication of *The Phenomenology of Spirit*, then what does this mean for the truthfulness of his idea of the spirit's evolution? The many parallels between certain

[36] Some might find it of interest that this stage is analogous to Kantian or critical idealism, asserting that "we cognize of things a priori only what we ourselves have put into them." I will leave this particular notion at that, reserving further discussion for philosophers.

contemporary disciplines and those of Hegel's time discussed in this chapter suggest that intellectual development is not exactly straight-forward.

A third interpretation of the title is based on a comment made by Slavoj Žižek in a section of *The Sublime Object of Ideology* where he analyzes Hegel's phrase in terms of representation. In his reading, the contradictory nature of the proposition itself – the reduction of the spirit to a fixed, dead object – provokes an unbearable sense of discord. The phrase, Žižek argues, is a prime example of a *speculative proposition*, and it is the failure of making sense of the discord inherent in it that finally leads to success in making palpable subjectivity itself: "we succeed in transmitting the dimension of subjectivity *by means of the failure itself*, through the radical insufficiency, through the absolute maladjustment of the predicate in relation to the subject" (2008 [1989]: 235 [italics in original]). In line with what I have already discussed, physiognomy was a kind of language, or system of signifiers. But as Žižek puts it, the discipline's final result is a complete failure precisely because the physiognomist could never actually find what he was looking for: "every signifying representation 'betrays' the subject; it perverts, deforms what it is supposed to reveal; there is no 'proper' signifier of the subject" (2008 [1989]: 235). In this sense, what the skull represents is pure negativity, the "impossibility of the signifying *representation* of the subject" (Žižek 2008 [1989]: 236 [italics in original]).

The proposition "the Spirit is a bone" for Žižek then, constitutes a movement from "the lack of the signifier into the signifier of the lack" (2008 [1989]: 237). To apply this interpretation to the artwork *Spirit is a Bone*, the title can be understood as a signifier of the lack of representational coherence between the images and their captions. Following Žižek's reasoning, this discord invites even deeper reflection on the representative nature of the portrait form itself. Furthermore, when the dramatic stripping down of the face into pure biometrics is compared to the work of Sander, it becomes clear exactly how communicative his photographs really are. This illuminates how Sander's approach to his portrait subjects was actually more dialectical than what the photographer's own way of talking about his work gave away. By literally giving face to an idea that is otherwise only thought or communicated abstractly, *Spirit is a Bone* invites the audience to approach the sense of discord it engenders as something material and of acute importance.

Conclusion

All sciences are rooted in creative interpretation, but some scientists are more aware of their own subjectivity than others. The ones presented in this chapter are clear examples of scientists who, in Gould's words, saw themselves as "apostles of objectivity." Moreover, physiognomists,

craniometrists and phrenologists did not really use their carefully gathered data to create *new* theories, but rather to confirm their preconceived a priori conclusions. Sander, although no scientist, was no exception to this rule. Just like Lavater and Broca, he saw in his own portraits what he had set out to depict, which was *en bloc* representations of German archetypes. Interestingly, however, his photographs have pleasing aesthetic qualities, strikingly expressive of individuality. Looking at them, one can experience layers and layers of meaning unfold. This stands in stark contrast to the portraits of *Spirit is a Bone* by Broomberg & Chanarin. Compared to Sander's *People of the 20th Century*, the subjects appear radically non-descript, which renders the captions almost paradoxical: unable to read anything meaningful about the sitters from the photographs, the viewer is inclined to consider them as a group rather than individuals. In this sense, the work becomes a critique of a system of surveillance – or a "way of seeing," as John Berger would have put it – more than anything else.

The contrast with Sander also serves to demonstrate that when a photograph is experienced as degrading towards the portrait subject, this effect cannot be attributed to the camera alone. The staging process, as well as the final print format are important factors in bringing forth this reflection. I find it especially interesting that only by looking at the installation in Hamburger Kunsthalle, I already had a sense that the subjects were not aware of being photographed. Whether this was a result of the abstractive processes of the AFR system used to make the portraits alone, or if it was rather an effect of careful artistic curation by Broomberg & Chanarin remains unanswered, but somewhere along the many-step process from caption to presentation, an air of inhumanity became tangible. The remaining traces of this digital process further raises questions about how this form of identification relates not only to the persons it is used against today, but also to historical predecessors. The discussion in this chapter makes it clear that in its quest for positive identification, contemporary facial recognition techniques can meaningfully be interpreted as the latest peak in this more than 200-year-old strive by local, national and international authorities towards order and control. This last point will be further developed in the following Chapter, where I will present a meta-analysis of all three artworks, further developing the most interesting finds.

Chapter 4

Dataismus: Disrupting Flow in the Age of Digital Reproduction

Introduction

Up until now, this thesis has treated ImageNet Roulette, *How do you see me?* and *Spirit is a Bone* as separate case studies. In the current chapter, the goal is to summarize the findings and discuss the works collectively. What can a meta-analysis of these three projects tell us of AFR and the use of photography in such technologies? To answer this, I have divided the chapter into two parts: the first will go through the general findings and discuss the most intriguing questions in more detail from an expanded historico-political perspective. As the reader is well aware by now, today's facial recognition systems clearly have a protohistory in nineteenth century positivist disciplines. My main goal in the first part will be to investigate this relation further in respect to all three art projects, not just *Spirit is a Bone*. I will focus specifically on the role played by the disciplinary systems in Europe and North America in the history of photography, expanding the discussion from Chapter 3. In this discussion, it becomes increasingly clear how, in addition to archives, statistical methods have played a central role in the medium's history – especially regarding the conception of "social types." Moreover, I will argue that it is no coincidence that these developments happened alongside the rise of the modern surveillance state. Can phrenology and contemporary surveillance systems both be understood as the products of national security concerns?

Zooming out from the material history of photography, the second part focuses on exploring what the works can tell us of digitality as a phenomenon in everyday- and institutional contexts alike. This particular discussion will center around conceptions of *flow* and other liquid metaphors often associated with the Internet and photography alike. In discussing this, I will ask whether the three art projects, ImageNet Roulette, *How do you see me?* and *Spirit is a Bone* can meaningfully be conceived as endeavors to de-naturalize the ideological foundations of AFR along with the apparent seamlessness characterizing our increasingly digitalized lives. As noted in Chapter 1, ImageNet Roulette might be understood as a form of activist project, or *artivism*; a term which could just as well be applied to the other two projects. Wondering if the works can perhaps be understood as *photomontages*, I will conclude by comparing them with Dadaist practice in the years following the First World War. In this discussion, I am especially interested in the aesthetic qualities that all three projects have in common: one is that they are presented *serially*; a quality which unquestionably affects how each single image is

encountered. A second is the fact that all four artists have appropriated existing technology, including its photographic basis, and re-assembled it in order to engender novel readings. The Dadaists sought to mystify the tendencies for which they blamed the destructions of war in order to oppose them. Can the works presented in this thesis similarly be understood as endeavors to contest contemporary identification schemes and arrest digital flow?

I

Taxonomies, archives, bias

In Chapter 1, it was established already in the reception analysis of the app how the biases inherent to ImageNet provoked the audience of *ImageNet Roulette* to the extent that the app went viral almost immediately after its release. Racist and misogynist slurs were especially prominent. In the related essay *Excavating AI*, Kate Crawford and Trevor Paglen criticized what they saw as unsupported epistemic ideas concerning the relationship between image and meaning informing the taxonomic structures informing AFR through the training databases. While any taxonomy is necessarily fueled by politics, the authors argued that this becomes all the more problematic when those issues are obfuscated through notions of mechanical objectivity and applied to categorize human beings. It is then that the taxonomies take a really dark turn by labeling people according to stigmatizing terms. Through a reading of that essay, it was established that the biases revealed by the roulette came from the taxonomic structure of the database in which the algorithm had been trained. This is why the duo were so intent on drawing attention to these so-called training image datasets. The exhibition *Training Humans* primarily served to affirm the general argument of the app and essay, while also outlining the evolution of the photographic portrait as used in facial recognition research.

Heather Dewey-Hagborg's *How do you see me?* seems to affirm that 2019 constituted a time of increased academic interest within the arts towards TIDs. The project was commissioned by the Photographer's Gallery in London as part of a broad digital program exploring the role and effects of training datasets. While I chose to focus on other aspects of the work in that chapter, the predominantly white appearance of the shapes recognized as faces by the algorithm in *How do you see me?* seems to support the argument that facial recognition algorithms are biased in relation to skin-color. As a result of the so-called black box problem, however, exactly what the shapes and colors signify to the algorithm remains a mystery.

Spirit is a Bone took up the same problems as those mentioned above but situated them in a specific (art)historic context. In this way, the work facilitated a reading of AFR technology as the latest progression in a centuries-old strive towards a method of positive identification.

Discussing this work in Chapter 3, I made it clear that early identification schemes building on physiognomic ideas began with studying the bodies of known criminals according to racialized, classed and gendered assumptions about what constitutes deviant looks and behaviors. As a rule, it can be said that those who were already stigmatized (POCs, migrants, the poor, etc.) were further subjected to institutionalized discrimination justified by the aura of scientific objectivity offered by the photographic apparatus. By appropriating the categories of August Sander, Broomberg and Chanarin made it clear how these issues also informed artistic photographic projects well into the 1900s. Although Sander himself arguably had an inclusive idea of the German people, his embrasure of the taxonomic scheme illustrates how the same positivist assumptions that served racist systems of discrimination well into the 20th century were prevalent across the political spectrum in Weimar Germany. Thus, the appropriation of Sander's image universe served to connect current AFR technology to precursors such as the racial science of the German National Socialist, which they used to legitimize the Holocaust. The South-African Book of Life used by the Apartheid regime has also been mentioned. In light of this, the fact that IBM turned to measuring *skull shape* when encountering problems regarding "fairness" in one of the company's facial recognition programs, as mentioned in Chapter 1, reads all the more disturbing.

At this point, the discussion presented in this thesis has demonstrated that not only is the history of photography closely intertwined with other technical and scholarly developments within European bureaucracies in the nineteenth and twentieth century, but also that these were largely designed to reproduce existing social and economic hierarchies – some more intentionally than others. What might still be unclear, though, is the wider political and technological contexts in which these developments took place. What questions were they designed to answer and what tasks were they intended to solve? To understand this, one needs to understand that phrenology and craniometry were all part of national surveillance systems. As has been pointed out by the sociologist Richard Jenkins, social life as always been monitored and regulated by surveillance in some form or other (gossip, word of mouth etc.). A defining characteristic of modernity, however, is the emergence of *formal, bureaucratic organization* as the standard way of organizing collective activities and enterprises (Jenkins 2012: 164). In order to connect the foregoing discussions to the context in which the scrutinized AFRs belong, I therefore need to take a step back and give a general presentation of the emergence of the modern surveillance state.

The Two-Faced Nature of Surveillance

It is not easy to pinpoint exactly when the modern surveillance state appeared, as the collection of demographic data has been a central part of state history for a long time. Simply put, however, one might say that it developed as a consequence of industrialization from the nineteenth century onwards. It was along with this development that states started to collect data about its population in a systematic, regulated and centralized fashion, most evidently in Britain, USA and continental Europe (Weller 2012: 58). A central aspect of the burgeoning national states was that they also offered some historically new privileges to their citizens: at least since Otto von Bismarck, first chancellor of the German Empire, introduced the first pensions and disability benefits in the 1880s, surveillance has been justified with the double need for welfare on the one side, and warfare on the other (Weller 2012: 59). In the beginning, people were asked openly for information regarding employment, family situation, personal economy, etc. This was done in order to determine who qualified for social services as well as for statistical purposes, which I will get back to later. In this way, domestic surveillance for the most part took place in an open manner, as people were inclined to submit to give up personal information in order to receive needed benefits.

By the end of the nineteenth century, more and more covert methods of surveillance were employed, often defended on grounds of national security. This tendency was further stimulated by technological developments in the following century, escalating during the Cold War period when the fear of spies and infiltrators rendered the need for secrecy to be perceived as more important than ever (Weller 2012: 61). Following the 9/11 terrorist attacks in 2001, national authorities' habit of covertly gathering information about its own population were taken to new extremes. As Toni Weller points out, it is of little practical importance whether a perceived threat is real or imagined, as even a hypothetical threat against the state can in principle lead to an experience of acute need for strict control and surveillance from its side (2012: 61). The result is that citizens submit their information whether willing or not. Today, I think it fair to say that it is very hard – if not impossible – to live in Western societies without giving away a considerable amount of personal information to public authorities, as refusing to do so would deprive one of even the most vital goods such as medical services.

It is not surprising that the period in question also overlaps with the rise of what Foucault calls *panopticism*, described in his book *Discipline and Punish* where he traces the way modern nation states gradually shifted focus from corporeal punishment to penalizing the "soul" as a means to form law-abiding subjects (Foucault 1999 [1975]). This is in line with what Alan Sekula has argued; namely that the invention of the modern criminal cannot be dissociated from

the construction of a law-abiding body (1986: 15). He also stresses how phrenology and related pathologizing disciplines appearing throughout the 1800s were part of a larger tendency to medicalize the study of the mind (Sekula 1986: 11).[37]

It could also be argued that more recent technological developments have amplified that same tendency: the collection, storing and systematization of data is now easier than ever, and achievable within short timeframes with little concern of geographical restraints. A crucial difference after the emergence of the Internet – and especially after social media have become normal platforms of communication – is that the users are usually implicit in the surveillance by willingly sharing data in order to consume what is shared by others. The most central difference between today's surveillance situation and that of earlier times, however, is the role played by private companies. While user data have always been part of commercial interests (think, for example, of the market for ad time on TV), we find ourselves today in a situation where private companies are in possession of so much specialized data that they in many ways have even wider and deeper knowledge of their users' life situation than what even the most intrusive state is capable of achieving. As surveillance scholar David Lyon puts it, "surveillance has spilled out of its old nation-state containers to become a feature of everyday life, at work, at home, at play, on the move" (2002b: 13). In the two decades since that proposition was written, its truth value has only become clearer.

In the analysis of *Spirit is a Bone*, I mentioned that the system FaceControl 3D was usually used for crowd control, typically located at entrances. This captures the essence of biometrics: more often than not, such technologies are being used precisely in border-situations to control *passage*. While I did not discuss this point in any detail in Chapter 3, it is nevertheless clear that the use-value of any biometric system lies in its ability to sort people; and what is the point of sorting if not to discriminate – either by submitting some to special institutions, or to exclude others from certain privileges? This can be in the form of differential access to geographic areas, which in the last couple of centuries have been increasingly defined in terms of nation state borders. As was also mentioned in Chapter 3, the development of ID documents occurred in line with stricter rules for cross-border travel and can thus be seen as artefacts of differential mobilities – and possibilities. Systems concerned with handling ID are thus

[37] In light of this, I wonder if not the history of AI can be seen in parallel to those developments: while early researchers such as Turing were primarily interested in axiomatizing logical rules, it was the second generation of AI researchers who in the 1950s came up with the idea that the brain (understood as a neurological system) was the most suitable model to emulate when building AI systems.

inherently discriminatory, in both senses of the word. This point is summarized by the surveillance scholars Torin Monahan and David M. Wood:

> Identification schemes try to fix the identities of known and unknown bodies so that they can be assessed and sorted either in real time or in advance, in an anticipatory way, so that screening at borders or checkpoints becomes largely perfunctory for elite (white) travelers, while it remains an anxiety-producing, unpredictable ordeal for (racialized) others. (…) elite travelers might enroll in pre-screening systems (or use dedicated toll roads in non-border settings), whereas on the other end of the spectrum, refugees seeking to relocate in Europe from North Africa and the Middle East are rigorously questioned about their backgrounds and then entered into various identification systems, some using iris scans, to assess their level of threat or need (2018b: 121).

As many surveillance scholars have also noted, the subject of identification has during the last decades increasingly been tied to the problem of immigration, rendering national borders areas of intense interest. This tendency spiked after the 9/11 terror attacks, which led to a spate of proposals internationally for new and enhanced ID card systems (Stalder & Lyon 2002).

Digital technology scholar, Kelly Gates, has written illuminatingly on how security agencies and politicians in the aftermath of 9/11 became increasingly obsessed with installing AFR systems in airports as a means to prevent further attacks. The promise of facial recognition, as she describes it, lay in the technology's potential to *individualize* the threat by targeting specifically identified "terrorist" individuals. Furthermore, "the effort to define it as a homeland security technology [in the US] also made use of an implicit classifying logic, including rhetorical moves that resuscitated antiquated notions of deviant facial types" (Gates 2011: 101). In other words, physiognomic notions of identity informed the post-9/11 security debate, presenting AFR technology as uniquely suited to identifying the individual faces of presumed terrorists. This development is very telling of how effective such rhetoric is, as AFR systems in these contexts are not even designed to classify faces according to racial typologies, but rather to look for specific suspects on terror lists etc. Thus, the symbolic authority of the technology in this context, Gates argues, "depended on the idea that it could in fact be used to identify a mythic class of demonic faces that had penetrated the national territory and the national imagination" (2011: 101). Gates describes this phenomenon as a "facialization" of terrorism – which was, paradoxically, often spoken in the same breath as claims about technical neutrality (2011: 101).

Focusing on the Netherlands, Karolina La Fors-Owczynik and Irma van der Ploeg describe how migrants and travelers are increasingly seen to pose a risk to the country, and that this problem is usually formulated in terms of managing *identity* (2016). Furthermore, their

research shows that despite increasingly complex methods for connecting immigrant bodies to their claimed identities, the various Dutch institutions involved in immigration control are still to some extent reliant on employees' gut-feelings (2016: 270). Of special relevance to the discussion in this chapter is how a central factor in the development of methods for risk-assessment in relation to migrants is "the presumption that risks are somehow observable" (2016: 270). What we see here is an echo of the early identification schemes discussed in Chapter 3, which also had as their primary goal to counteract fraud – a novel threat in newly urbanized societies. In relation to national borders, however, the problem is less one of tracking down and punishing criminals, but rather to *keep out* unwanted immigrants in the first place. This can, for example, be seen in the Schengen Information System (the main EU police database) where illegal immigration constitutes an overwhelmingly large share of the system's usage (Aas 2018: 138). I argue that this, although not explicitly expressed, should be understood as an underlying theme in the artworks presented in this thesis, as AFR systems cannot be understood independently of the problems they are designed to solve. In this, the works fit into a broader tendency within the art world. As the art historian David Hopkins has noted, the War on Terror has served as a backdrop for contemporary art during the last two decades, which can be seen, amongst other things, in a surge of works thematizing asylum seekers. As he puts it, "borders have become highly potent symbols of the current age" (Hopkins 2018: 253).

As should be clear by now, any surveillance system can be understood as a practice of social sorting; an approach which has been strongly argued for by David Lyon. As he puts it, "surveillance is not itself sinister any more than discrimination is itself damaging," which is why a focus on the organization of personal data, as well as the social and economic circumstances surrounding such practices is paramount (Lyon 2002a: 2). Discussion of social sorting is very important because it is central to the formation of people's life-chances and decides their opportunities – in particular through the allocation of rewards and punishments. As was briefly noted in Chapter 3, biometric technologies serve to affect already othered communities the most consistently because of how they are "marked by the persistence of problematic assumptions about difference" (Magnet 2018: 119). As others have noted, this is true irrelevant of whether the members of a particular category themselves know about being profiled (Jenkins 2012: 161).

Given the above implications, society as a whole should be very concerned with issues of surveillance. To repeat Louise Wolthers' argument discussed in Chapter 1, the responsibility to do so weighs extra heavily on theorists and artists in privileged positions. In my own reading

of the works, it is just this responsibility that Paglen, Crawford, Dewey-Hagborg, Broomberg and Chanarin have taken upon themselves through their artistic projects.

The Ultimate Rabbit Hole

While the most obvious difference between digital and analogue surveillance is *quantitative* (computer hard drives can accumulate and store far more information than analogue systems), the more relevant difference in this particular discussion is how digital systems enable automatization of these processes. This allows for unimaginable amounts of data to be processed at increasing speeds surpassing what any group of humans could ever dream of achieving manually. From this perspective, AFR systems seem to answer a nineteenth century dream: as Alan Sekula has noted, the problem of processing was precisely why Bertillon's system eventually fell out of use. After a couple of decades, the police archives were so massive that it became practically impossible to find the relevant documents within a reasonable timeframe – especially since arrestants often gave false information. For Bertillon, this growth constituted a problem of organization: after a long struggle to establish efficient identificatory methods, he therefore changed his focus to coming up with methods for organizing and searching the archives. In Sekula's words, he became a "compulsive organizer," but eventually had to give up because the task was simply too demanding (1986). In this, Bertillon's attitude paralleled current developments.

Intuitively, it would seem that Bertillon's prayers were answered with the pattern discrimination algorithms of today. As Kate Crawford has noted several places in her academic work, however, another problem tends to arise, informing intelligence work across the globe like a disease. She describes a particular kind of anxiety driving the endless collection of ever more data. In some of the leaked Snowden documents, it is clear that even people within the NSA themselves have expressed concern in relation to what they call *summit fever* – a play on the term used to describe the phenomenon of mountaineers' compulsion to reach the summit of a mountain at all costs. This, Crawford explains, in a surveillance context denotes the fear of missing crucial information or not seeing the right connections akin to Jacques Derrida's concept *archive fever* (2016). The problem with this, as Derrida made clear, is that the archive produces just as much as it records and can only ever offer a particular kind of reconstruction, a partial view (1996). If we take the twin anxieties of the surveillants and the surveilled and push them to their natural extension, Crawford brilliantly describes how we reach an epistemological end point: "on one hand, the fear that there can never be enough data, and on

the other, the fear that one is standing out in the data." These fears, she argues "reinforce each other in a feedback loop, becoming stronger with each turn of the ratchet" (Crawford 2014).

The above is in line with what Lyon has also argued, that surveillance is growing not merely because new devices become available, but also that "devices are sought because of the increasing number of perceived and actual risks and the desire more completely to manage populations – whether those populations are citizens, employees, or consumers" (2002b: 20). Thus, it appears that the assembling of data and construction of digital archives is the ultimate rabbit hole (remember that ImageNet's original mission was to "map out the entire world of objects"). This reminds me of an observation made by Alan Sekula in his influential essay *The Body and the Archive*: "since physiognomy and phrenology were comparative, taxonomic disciplines, they sought to *encompass an entire range of human diversity*. In this respect, these disciplines were instrumental in constructing the very archive they claimed to interpret" (1986: 12 [italics added]). It is clear, then, that the taxonomic practices discussed in this thesis echo the physiognomist and phrenologist practices not only in their particular interest in the human skull, but also in the belief that truth lies in quantity. The myth that continues to inform this drive is the idea that if you only look long and close enough, eventually you will find the truth.

Pattern Discrimination and Determinism

International relations scholar Mark Salter has noted that the marking of some individuals as "high risk," more than anything indicates a *lack* of precise knowledge, suggesting only suspicion. This, in turn, leads to a spiral of insecurity within the bureaucratic structures of control (2018: 130). Salter's point echoes one of Sekula's many clever observations, which is that the early instrumental issues of what he calls *photographic realism* "were systematized on the basis of an acute recognition of the *inadequacies* and limitations of ordinary visual empiricism" (1986: 18). This, he claimed, was exactly why they needed to ground the photographic evidence in more abstract *statistical* methods. It seems then, that not only has photography always been related to physiognomy, but that this was also the case for mathematical statistics. This renders a complete analysis of the medium's role in AFR incomplete without further discussing this issue.

While the taxonomic systems are produced by researchers, the specific biases inherent to the coupling of certain categories with certain images is a result of the collective work of remote laborers who are often hired on crowdsourcing platforms such as Amazon Mechanical Turk. However, when the learning process itself is left to opaque abstractive processes, new issues arise. So, what happens when these processes are increasingly being automated? As

Monahan and Wood have put it, "state identification projects offer the most glaring examples of how discrimination may become encoded in – and reproduced through – abstraction" (2018a: 94). As I have shown, this takes on new dimensions in relation to photography.

Sekula was the first to systematically argue that photography from its very invention became ingrained in the wider context of demographic regulation and so-called social statistics; a science invented by the Belgian astronomer and statistician Adolphe Quetelet to handle social phenomena mathematically. While Bertillon was more of a clerk interested in streamlining and improving the efficiency of police work, Lombroso, on his side, was primarily interested in social types. In this, his basic idea can be traced to Quatelet, who in 1835 introduced the idea of a *composite character* derived from large aggregates of social data (Sekula 1986: 20). With conceptual tools borrowed from astronomy and probability theory, he observed that variants of social data, when derived from large numbers, often fell into a pattern corresponding to a bell-curve. This led him to formulate the idea of "social physics," which saw the social norm as a force of "gravity" towards which individuals were drawn (Sekula 1986: 21-22). By the end of the nineteenth century, this model, which Sekula refers to as an "organismic model of a visible social field," was in crisis. The demise of the organismic model was, according to him, largely due to Gabriel Tarde, a criminologist with background in law, who was critical towards the determinism inherent in that theory. He and his colleagues in the so-called French school of criminology stressed the importance of environmental factors in determining criminal behavior, which they often related to bad environmental effects of urbanism (Sekula 1986: 37).

By the end of the century the discipline had drifted towards this school, gradually replacing Lombroso's Italian variant. This, however, did not bring an end to statistical treatment of photographic portraits. As was argued in Chapter 3, AFR is essentially based on statistical calculations. This has been discussed interestingly by the artist-philosopher (and so much more) Hito Steyerl, who has written an excellent essay on pattern recognition. As she points out, it is necessary to come up with systems to extract meaning in order for the unimaginable quantities of data to be of any use. This is where pattern recognition comes in handy, as it offers a means of navigation. But this, Steyerl stresses, involves differentiating between "meaningful" data and mere "noise," which leads to the problem of determining what should be considered *meaningful*. As the philosopher Jacques Rancière has argued, what is defined as noise is always political: whose voices are considered speech (and listened to), and whose voices are filtered out as mere noise is determined by existing social hierarchies in what he refers to as the *distribution of the sensible* (Rancière 2012 [2008]). This distribution, for Rancière, is a

constituent part of democracy. Drawing on this, Steyerl argues that "dividing signal and noise means not only to 'filter' patterns but also to create them in the first place" (2019: 3).

The data-equivalent of noise is *dirty data*. Unsurprisingly, Steyerl describes how these are singled out on the basis of existing conceptions of reality. The example she gives is a story of a team of analysts studying the demographic data of guests at a luxury hotel chain: the data indicated that whole groups of 17-year-olds from wealthy Middle Eastern countries were staying at hotels all over the world, but because the idea of rich Arab teenagers appeared too contrary to the analysts, the data were dismissed as dirty, only to be verified as true at a later point (Steyerl 2019: 5-6). This is an example of how data about the world is typically adjusted to existing world views. As Steyerl puts it, "probability enters truth production on an extensive scale with the unsurprising effect that the patterns supposed to be uncovered in massive data correspond to some degree with the patterns that are already assumed to be there" (2019: 6).

Steyerl's discussion of dirty data echoes the way Broca reacted when he encountered data contradicting his idea of Europeans as superior to all others. As I briefly mentioned in Chapter 3, he at one point found that some "yellow people" scored better than the average European on brain size. While this forced Broca to admit that "a lowly race" *might* after all have a big brain (which would discredit his whole theory), the problem spurred some impressive cognitive acrobatics by the craniometrist. I cite Gould here in more detail than before, as the analogy to the above discussion is especially clear the way he tells the story:

> But Broca felt that he could salvage much of value from his crude measure of overall brain size. It may fail at the upper end because some inferior groups have big brains, but it works at the lower end because small brains belong exclusively to people of low intelligence. Broca continued:
>
> But this does not destroy the value of small brain size as a mark of inferiority. The table shows that West African blacks have a cranial capacity about 100 cc less than that of European races. To this figure, we may add the following: Caffirs, Nubians, Tasmanians, Hottentots, Australians. These examples are sufficient to prove that if the volume of the brain does not play a decisive role in the intellectual ranking of races, it nevertheless has a very real importance.
>
> An unbeatable argument. Deny it at one end where conclusions are uncongenial; affirm it by the same criterion at the other. Broca did not fudge numbers; he merely selected among them or interpreted his way around them to favored conclusions (1997: 119).

The situation described above is an extreme example of motivated interpretation if there ever was one. Determined to arrive at his a priori conclusion at all cost, Broca merely discredited the numbers that would disprove his theory as dirty data.

An important difference between Broca and machine learning, however, is that the craniometrist's inconsistent musings were actually written down and published. This, and a long list of other outrageous examples of unscientific methods, was precisely how craniometry was discredited. With machine learning algorithms, however, the black box problem renders such scrutiny all but impossible. That the procedures involved in the training process are non-traceable and non-reproducible is precisely why the field is being criticized for lacking scientific rigor. One of the most basic aspects of scientific theory is that its explanations should be public and intersubjectively repeatable; meaning that another scientist in theory should be able to conduct the same research and arrive at the same results. This obviously cannot be said of neither machine learning nor the Victorian pseudo-scientists. While one might be inclined to laugh at the stupidity of the latter, I argue that in actually publishing the data and formulating their line of reasoning, Broca and his colleagues were in fact *more* scientific than most machine learning research is at present. Here, it should be stressed that unlike the systems discussed in this thesis, the algorithms and taxonomies informing most AFR systems are *not* accessible to the public. The ones discussed here are therefore potentially the least problematic ones.

To be fair, computer scientists, when criticized, are usually careful to stress that their algorithms should only be used as guiding *tools* and not to form definite conclusions. In practice, the latter option is nevertheless how they are being employed. One such example is the story of Robert Julian-Borshak Williams, who was arrested by Michigan police in 2020 after an AFR system had misrecognized him as a shoplifter. When he came to the police station, the officers showed him the security footage and asked if it was him. It was not: "you think all black men look alike?" Williams answered. Realizing the mistake, one of the officers answered: "I guess the computer got it wrong" (Hill 2020). This story is almost tragicomic in its banality. While Williams was released shortly after the interrogation, he had already spent thirty hours in arrest, after first being handcuffed in front of his young children. That a quick comparison of the photos with the man so easily revealed the error only serves to illustrate the degree of faith put in AFR. This makes the story an example of determinism in practice: as algorithms are clearly biased along racial lines, it perpetuates existing stereotypes. Here, it is the black man as "offender" – just like Mr. Kima mentioned in Chapter 1 was labeled by *ImageNet Roulette*.

The idea that determinism is inherent to taxonomic systems is explicitly mentioned in the essay *The Bone Cannot Lie*, where Eyal Weizman discusses *Spirit is a Bone* with Adam Broomberg and Oliver Chanarin. Here, he reflects on the temporal dimension shared between all facial recognition technologies: through probabilistic methods, they seek to detect dangerous subjects already *before* a crime has taken place. Interestingly, Weizman also connects this

technology to 9/11, arguing that systemic determinism has been amplified in the aftermath of the terror attacks. He points to out how phrenology was the first science that sought to "peer into the future": by establishing what types of people were likely to engage in criminal behavior, preemptive measures could be initiated. The same, Weizman argues, is true for AFR due to the special nature of terrorism:

> Any beginner terrorist mastermind knows that "important" operations must employ operators without any criminal or terrorist track record. These are crimes perpetrated by people that have been innocent before they took place and dead immediately after the event. The transition between innocence and death is so short, almost instantaneous, so the states perceive their task as needing to look into the future because the past cannot be mined and the present is too short to tackle (Broomberg & Chanarin 2015: 225).

Because many perpetrators die in the act, going after them after the event would make no sense. The only option left to combat terrorism is thus to calculate the future mathematically and to observe any potential risk closely. This brings Weizman to describe the archive created for *Spirit is a Bone* as an "offshoot on the long war on terror" (Broomberg & Chanarin 2015: 221).

In the case of terrorism, the future has become the domain of statistics, and the task facing intelligence agencies is to search for patterns that might give off who is a threat. It is not for nothing that the laws concerning terrorism are in many countries amongst the strictest. In 2002, for example, the Norwegian Storting passed *terrorparagrafen* ("the terror clause"), which allows for prosecution on the basis of *planned* as well executed acts of terror. It opens for imprisonment up to 21 years (with potentially longer detention), which is the country's maximum penalty (e.g. Nilsen & Olaussen 2012). The clause is designed precisely to be able to punish *imminent threats*; a term usually associated with the War on Terror. In this way, present-day surveillance systems, in their ontological relation to the surveilled subjects, are not so different from the physiognomic idea of deceptive outer appearance. The archive, in this scenario, becomes the body of knowledge through which suspicious patterns can be determined by the algorithm – be it a collection of photographs, personal data, or both. In light of this, I find it interesting how Walter Benjamin, in his 1936 essay *The Work of Art in the Age of Mechanical Reproduction*, commented (with dismay) on how the visual field in his time was increasingly affected by the "growing importance of statistics." As he saw it, this development was of "unbounded consequence not only for thought but also for the way we see things" (2008 [1936]: 10). In light of the foregoing discussion, Benjamin's diagnosis of the visual field appears especially clear-sighted, as this has only accelerated with networked digital photography.

Before rounding of this first part of the chapter, I will discuss one last thing, which is that pattern recognition is also a human ability, and not specific to computers. This has been noted by no other than the iconic whistleblower Edward Snowden:

> Humans are by nature pattern-recognition machines. We search for meaning, whether in the circumstances of our lives or on the surface of toast. Like computers, we are incapable of producing truly random values over time. This is why our phone's password is closer to "2846" or "kittymittens2" than "DH?zM#XUAgoIj/1;R16I<<rK7'6k" (2016: 121).

In relation to this, Steyerl's previously discussed essay is again relevant. Here, she also takes issue with a Google experiment conducted some years ago where the researchers applied image recognition algorithms on sheer random noise: there was nothing to recognize, as the data were completely random, representing nothing. Interestingly, the algorithms ended up recognizing things anyway; shapes "started emerging where combinations of the shapes and animals the networks had been taught to 'see' earlier on" (Steyerl 2019: 9). The result was a clear display of over-identification. This phenomenon of detecting patterns in images where none exist is referred to by Google as *inceptionism*. *How do you see me?* is a good example of this phenomenon, as it displayed how the algorithm recognized faces that were not actually there. To revoke Dewey-Hagborg's question of what this might mean, it is evident that people might be profiled or prosecuted on the basis of erroneous identification – a risk which is seemingly higher for ethnic minorities.

This is not to say that only computers are guilty of over-recognition, as the phenomenon of apophenia clearly demonstrates. This is the tendency to perceive meaningful connections between unrelated things and is typically exemplified by seeing shapes in clouds. Steyerl uses this phenomenon to draw parallels between algorithmic pattern recognition and ancient peoples who saw shapes in the stars, such as a crab – which they called Cancer. While one could laugh at this naivety, Steyerl stresses that had they not projected fictional figures onto the cosmos, later people would probably not have discovered the fundamental movements of the solar system. Then again, these patterns would not have been discovered if astronomers had not at some point admitted that the shapes which astrologers had originally identified were *not really there.* As *How do you see me?* demonstrates, algorithms treat the shapes as if they *are* really there. Steyerl summarizes this problem and its particular kind of determinism amusingly:

> Computer vision still seems to be in the phase where it thinks that there really are crabs in space and that the patterns emerging from the cosmos of data is actually reality. Software engineers like saying about computers: garbage in, garbage out. In

> deviationist computer vision, let's rephrase this as: crab in, crap out. Let's see faces in clouds while we're at it! (Steyerl 2019: 15)

This perfectly captures the problem at hand, and also why we should be extremely careful with letting insufficiently tested AFR – or any other kind of machine vision systems – influence our surroundings. If not, we might expect a future that stinks.

It is evident that both modern and historic methods for facial recognition are based on statistical models of reality, but when computers are brought into the picture, the social power of information is reinforced. As David Lyon has put it, the data doubles "start to flow as electrical impulses, and are vulnerable to alteration, addition, merging, and loss as they travel" (2002b: 22). What's more, the ongoing life of the data doubles depends upon complex information infrastructures. On the one hand, Lyon argues, this may help to democratize the information, but on the other, it may equally lead to tyrannies through "the quiet victories of infrastructure builders inscribing their politics into the systems" (2002b: 22). Given the inherently heterogenous nature of images, this is paradoxical. As image theorists Daniel Rubinstein and Katrina Sluis have noted, images do not have a limit or an end, preventing them from "becoming pre-determined, or conditioned on some kind of moral imperative, logical necessity or instrumental indexicality" (Rubinstein & Sluis 2013: 34). The various chapters in this thesis have demonstrated that this is exactly how photographic images are nevertheless employed in AFR technologies. It is this paradox that Bruno Latour and his colleagues point to when they, in a discussion of Tarde, state that "the whole is always smaller than its parts" (2012). Statistical representations, or data doubles, are always reduced and less complex variants of the persons they are supposed to represent and should therefore be treated as such.

II

Digital Dualism

From the above discussion, it is evident that already from the outset, conceptions of identity have been central to the development of photography as a tool in facial. In a networked context, the matter is further complicated because of some widespread ideas concerning the nature of digital phenomena more generally. Christiane Paul, an authority in the field of digital art, has commented that "virtual existence suggests the opposite of a unified, individual body – multiple selves inhabiting mediated realities" (2015: 165). In her book on the topic, Paul relates this to how online identities allow for an experience of being simultaneously in various spaces and contexts, facilitating an experience of "reproduction of the self without body." The complex interplay between "virtual" and "physical" existence, she argues, affects our understanding of

the relationship between body and identity (2015: 165). In light of this, it is not surprising that identity has become an increasingly common theme in digital art, putting to question the materiality of online experience. It seems that people are generally becoming more aware of how engaging with digital tools is also an embodied practice, as "current interface standardization has led to an overall restraining mechanism for the human body, which is forced to conform through computer and monitor" (Paul 2015: 170). Because we have to physically adjust to our digital gadgets and tools in both everyday and professional settings, there is a tension between the actual performance of using these, and the experienced disembodiment in networked communication as detached from the realm of the primary senses – as if there is an invisible barrier rendering everything happening "inside" the screen ontologically distinct from what's going on outside of its material surface. Especially within the realm of aesthetics, it has become clear that this tension "can not be constructed as a choice of either/or but rather has to be understood as a reality of both/and" (Paul 2015: 170). The current climate thus stands in sharp contrast to the early days of computer art, when it was common for artists to seek means of transcending the supposed opposition between mind and machine (Popper 1993: 121).

At least since the early 2010s, media scholars have been aware that that the idea of *digital dualism* is a fallacy. I borrow this term from Nathan Jurgenson, who describes how the distinction is not only false, but that the perceived binary also wrongfully privileges the "physical" world as *more* "real" than the "virtual" one, when the two realms are in fact increasingly enmeshed. As he points out, "our Facebook profiles reflect who we know and what we do offline, and our offline lives are impacted by what happens on Facebook" (Jurgenson 2011). In the book *Glitch Feminism*, writer and curator Legacy Russel builds on this discussion. She points to how the abbreviation "IRL" (In Real Life), which is common slang in online chats to denote life "offline," reflects the prevalence of this false binary. Instead, she proposes to substitute it with "AFK" (Away From Keyboard) – also a common expression online – because it supposedly better embraces the relation between online and offline experience. In Russel's view, AFK "signifies a more continuous progression of the self, one that does not end when the user steps away from the computer but rather moves forward in society away from the keyboard" (2020: 30-31).

While I am in complete agreement that the vocabulary currently available falls short of grasping the complexity of the relationship between online and offline identity, I do not think that AFK sufficiently does so either. The primary reason for this is that it privileges the computer with its keyboard as the primary site for online activity. Secondly, I question whether it is meaningful to describe life offline as "away" from our internet-connected devises. Do not

most people in the rich parts of the world nearly always keep a smartphone on them? In my experience, the majority is *constantly online*, always available for instant communication. While I cannot offer a more precise vocabulary at this point, I do believe in the possibility that this can develop naturally if digital dualism is continually contested. This is precisely what I experienced ImageNet Roulette, *How do you see me?* and *Spirit is a Bone* to do. Interestingly, the fallacy of digital dualism became especially clear in the gallery space of both *Training Humans* and in Hamburger Kunsthalle where I first encountered *Spirit is a Bone*. As I described in Chapter 1 and 3, these works facilitated a sense of conflation between the digital space in the artworks and the gallery space where I was standing. This made me question the notion of digital and physical space as separate. Moreover, the works also opposed the distinction between the private and public realms. This indicates that traditional offline galleries still possess radical potential in that they offer spaces where different modalities of contemporary life can be brought together and made perceptible.

While all three art projects spurred questions concerning the materiality of digital phenomena, I wish to dwell on how they applied some curious visual aesthetics, which themselves reflected some misconceived notions of digitality. I am thinking here of the use of neon green squares to indicate faces in both *ImageNet Roulette* and *How do you see me?*, evoking early conceptions of online space derived from William Gibson's novel *Neuromancer* (1984). It was here that the term "cyberspace" was brought into the world of popular culture, described as an immaterial "nonspace" separated from the physical world; a kind of collective hallucination by millions of people projected mentally as lines of strong light with a clear prevalence of neon colors. An example where the protagonist, Case, is performing a hacking operation is illustrative of this: "he punched himself through and found an infinite blue space ranged with color-coded spheres strung on a tight grid of pale blue neon" (Gibson 1984: 67). By the mid-1990s, the term cyberspace was solidly established in popular media, and so was the prefix "cyber" to denote anything internet related. Some commentators nevertheless argue that it has since worn out of fashion due to the changed nature of Internet 2.0, where simultaneity is no longer necessary for sharing and socializing online (Ryan 2014: 120). In light of how effective the use of neon green was for situating the artworks in a discourse on digitality, however, these visual ideas of cyberspace might not be completely out of fashion after all.

Spirit is a Bone, on its side, did not employ a similar aesthetic, but the installation version could nevertheless be said to play on the same notions of the Internet as a non-space. I am here referring to how the black background made the faces appear to "float" in an immaterial space. A tension can thus be located in all three works: in order to situate themselves in a

discourse to critique digital tools, they play on the misconceived visual language which itself serves to obfuscate the material reality of digital phenomena. In the following, I will discuss in more detail the artistic methods employed to bringing forth such reflections, but first, I wish to consider another linguistic aspect of the discussion surrounding digital phenomena and photography in particular; namely how these are often described in terms of liquidity.

Going Against the Flow

No man ever steps into the same river twice. We have all heard this expression by Heraclitus, but after interaction with the works discussed in this thesis, new dimensions may be added to the ancient observation. It is apparent that digital phenomena are often referred to in terms of liquid metaphors, such as *flow* or *stream*. Throughout this thesis I have repeatedly done so myself. One could also add expressions such as "surfing the Internet." Interestingly, the same is true for photography, a tendency which has been discussed by photography theorist Michelle Henning. Building on the work of Jonathan Crary, she relates this to how photography and other optical devices were developed in tandem with new forms of travel and mechanized industry throughout the 1800s. They thus "participated in the emergence of a new kind of optical experience that was mobile and detachable" (Henning 2018: 128). Not only did the telegraph arrive shortly after the camera, but the period also saw electricity installed in cities across the world. The nineteenth century as a whole was thus affected by an experience of collapsing of space and time through instantaneity and speed. "From its inception," Henning argues, "photography facilitated the movement if images in various ways, but this movement was hugely accelerated by the advent if electronically transmittable images" (2018: 131). The latter had been technically possible with the invention of telephotography at the turn of the century.

This development was only accelerated by the advent of digital photography. A curious consequence of this, as Henning describes it, is a recent shift in focus within photographic scholarship from the *singular* image and rather to *series*, repetitions, or sequences, as this supposedly better represents the nature of how photography is made and distributed. This is relevant to my discussion, as this is exactly how ImageNet Roulette, *How do you see me?* and *Spirit is a Bone* were presented. It is my understanding that this offers a reading of the images as representations of a *practice*, a phenomenon. Encountering the images serially, it becomes clear that the use of liquid metaphors is misleading – at worst completely obfuscating the actual materiality of photography and other digital phenomena (hardware, servers, cables, etc.). For me, use of the term "seamlessness" to describe design and user experience is what best captures the endeavor to make the materiality of gadgets appear to disappear. As one commentator puts

it, “at its diaphanous best, virtuality vanishes into sheer transparency. Once software flows smoothly throughout the contours of human gestures, needs, and desires, there is a feeling of unmediated activity” (Heim 2014: 515). Perhaps this was what informed my intuitive connection to the Internet 2.0 when entering the second floor of the gallery in *Training Humans*: in that the snapshots were distributed without gaps as a continuous flow – even the TV screens were placed amidst them – the interconnectedness implied by this curatorial choice reflected new and “seamless” interfaces as opposed to old, less smooth ones.

Before I move on, it is important to note that liquid “cyberspeak” is not new with the dominance of Silicon Valley tech-companies, but – like so much discussed in this thesis – dates back to the Victorian era when this was not just metaphorical jargon. When photography was invented, Henning points out, describing the medium in liquid terms was technically correct in that development was indeed a chemical process. The same goes for electricity insofar as analogue signals were continuous: the liquid metaphors were used with the intention of easily explaining the new technologies to lay-persons (2018: 135). This could even be said to hold true of a TV broadcast in that it was at least a continuous emission of signals. When it comes to the “streaming” of TV series, however, liquid terms are misleading euphemisms, as digital media “operate in bits, as discrete pieces of data, which are transmitted across the network in non-continuous fashion” (Henning 2018: 133).

At this point, it is important to note that liquid metaphors are commonly associated with economy as well. This has reached new dimensions in a market which is itself increasingly digitized, rendering the material foundations of monetary value even less obvious than previously. This general conception of state of affairs is exploited by tech companies, Henning argues, who are deliberately using terms such as flow or “clouds” to convince users of “seamless efficiency” (2018: 146). As opposed to this, it should be added, our laptops and smartphones are not seamless at all: they are based on asymmetric global distribution of capital and opportunities, exploiting unfair labor conditions in the global south – not to mention the environmental effects of producing and distributing the necessary materials (e.g. Byrne & Hudson-Edwards 2018). The use of underpaid Turkers fits right into this narrative.

When media theorists speak of “image flow,” on their side, Henning argues that they are referring to how media are *experienced* as “an overwhelming diet of rapid, endless and instant imagery” (2018: 146). The consequence, as she sees it, is that image flow is associated with the impossibility of fixing anything at all (2018: 146). In temporal terms, this is paradoxical in light of how photography has historically been understood as a means to *arrest* flow – to seize the *moment* from the *flow of time*. So how, then, are we to combat this

misconception in a way that enables a constructive approach to networked digital photography? This is where the work of Paglen, Dewey-Hagborg and Broomberg & Chanarin feel like a breath of fresh air, making perceptible the incongruous relationship between virtuality and the language surrounding everything digital. As the preceding chapters have demonstrated, many people – me included – instinctively reacted with suspicion towards the data double visualizations. May this be because the literal objectification which the works made visible appeared as a violation of sorts? If so, then the mystification facilitated by the works is a potent political tool. In a broader perspective, the works can even be said to renegotiate the artistic genre of portraiture more generally.

In his book *Portraiture*, art historian Richard Brilliant suggests that at the heart of the genre as an art form lies the fact that it invariably alludes to an individual human being existing outside of the work. This "vital relationship between the portrait and its object of representation", he argues, "directly reflects the social dimension of human life as a field of action among persons" (1991: 8). By employing AFR technology to produce portraits, however, this supposed social dimension is put to question. The images presented in the three projects contest the notion of portraiture as a "field of action amongst persons." One might also add that the works, by recontextualizing the AFR systems into the social sphere, facilitate a relation towards the technology where it is held morally accountable according to the same standards *as if* it were a person. This is especially true of *Spirit is a Bone*, in relation to which I had an immediate hunch of the non-reciprocal relationship between the portrait subjects and the photographer.

In this respect, it is relevant to mention that when French police first started to document arrestants photographically, this was experienced as a gross power abuse; an assault on the individual's integrity (Mortensen 2012: 77-78). With time, however, the ID photograph has been naturalized and come to be seen as an almost neutral form of control.[38] In light of this, it is pertinent to ask whether the art projects can be seen as endeavors to re-mystify the act of photographic documentation. Reflecting on this, I will compare them to Dadaist art, which

[38] Recent events indicate that this is true in Norway as well. I am thinking in particular of the app Smittestopp, developed by the Norwegian Institute of Public Health in order to prevent the spread of Covid-19. The app led to much controversy due to privacy concerns already before its launch, and only a small percentage of Norwegians actually downloaded it, rendering it practically useless (Dagbladet 2020). In comparison, I know of no recent public discussion of the ethical status of passport photography or biometric imaging at home, while Norwegian newspapers regularly write critically about the use of AFR in, for example, Russia and China (e.g. Almås 2019). Thinking of how ingrained biometric technologies have become in air-travel across Europe during the last couple of years, the contrast is striking. True, some Norwegian newspapers have discussed this issue to some extent, but the gradual development has, according to my observations, been normalized without further ado.

employed various artistic strategies to scrutinize what the artists viewed as the destructive ideas for which they blamed the First World War. I am thinking especially of Dadaist *photomontage*. By reading ImageNet Roulette, *How do you see me?* and *Spirit is a Bone* from this perspective, I intend to make clear the political potential in this particular artform.

Data or Dada?

Just like Dada, data by itself has no fixed meaning. Therefore, it can in principle mean *anything.* It is clear from the analysis in this thesis that in image recognition systems, a bouquet of biases usually leads to extraction of stereotypes from the endless possibilities which lie in *anything*. It seems that big-data fundamentalists are of the opinion that more data is inherently better, or closer to the truth. Discussing this, Kate Crawford has pointed out how there exists no point in this theology at which enough is enough. The tendency of compulsive data-gathering, she argues, is epistemology taken to its limit. As she puts it, "the affective residue from this experiment is the Janus-faced anxiety that is heavy in the air, and it leaves us with an open question: How might we find a radical potential in the surveillant anxieties of the big-data era?" (2014). This question might be exactly what the art projects presented in this thesis intend to answer.

From an aesthetic perspective, it seems that Paglen, Dewey-Hagborg and Broomberg & Chanarin have all approached the above question by revoking methods that are in many ways reminiscent of Dadaist photomontage. I am here operating with a broad definition of the term, including any artistic expression that uses photography as its primary imagery – either cut up or reassembled in order to produce new meanings. According to this definition, all photography could in principle be considered photomontage (as they from the get-go were commonly retouched). To distinguish photomontage from other merely edited photographs, I therefore stress the importance of the aesthetic expression to *deliberately* appear as re-assembled or re-contextualized. From this definition, it is clear that all three of the works at hand are examples of photomontage in that they make apparent how the imagery is taken from another context – that of AFR and surveillance.

Dadaism as an artistic movement first developed in Zürich in the mid-1910s as a response to the ongoing world war, but it is the Berlin group who are famous for their use of photomontage. For the German Dadaists – who belonged politically to the radical left – the main enemy was bourgeois sensibility. Reading the many manifestos produced by the Dadaists, it is clear, however, that the practitioners were extremely concerned with opposing *anything* that reeked of moralism, objectivism, positivism, logical systems, or anything else indicative

of "ultimate truth". This is perhaps most obvious in Tristan Tzara's *Dada Manifesto* of 1918, which is peppered with polemical utterances such as "I detest greasy objectivity, and harmony, the science that finds everything in order" (2011 [1918]: 141). Probably as a consequence of the extreme devastations the city suffered during the war, The Berlin Dada fraction were more explicit in their political stance than their colleagues abroad, breaking with the ironical treatment of war and politics exemplified by Tzara and the other originators of the movement in Zürich.[39] This was especially visible in the aesthetic expression of their work: while the Swiss Dadaists were radically iconoclastic, more focused on poetry and performance than on the visual arts, the German group made explicit use of the photographic medium's rhetorical possibilities. Common to all of the Dadaists, however, was that they opposed Art with a capital A, which they repeatedly declared dead while defining themselves as *anti-artists*.

The latter point is evident in the term photomontage itself, which, as art historian Dawn Ades has explained, was chosen precisely because it implied that the photo*monteur* was a *mechanic* rather than artist; situating him clearly within the working class and its struggle (1986: 12). Furthermore, the photograph, as some later explained, belonged to the technological world of mass communication and mechanical reproduction and therefore perfectly suited the purpose of integrating objects from the world of industry into that of artistic production (Ades 1986: 13). But how is this relevant to the general discussion of this thesis? In my understanding, there are several similarities between the works discussed here and Dadaist photomontage. Primary amongst these is the way they both appropriate existing photographs in order to make evident the epistemological and ideological framework informing the political reality in which they were made. Next, there is also a parallel between the context of the First World War and the current War on Terror in that they are both inherently international in scope and to a large degree concerned with national borders.

In his short essay titled *Political uses of Photomontage*, John Berger explains the political potential in photomontage in the following way: using photographs as its material, the montage keeps its familiar photographic appearance. Even if cut up and reassembled, it is clear that "we are still looking first at *things* and only secondly at symbols" (2013 [1969]: 30). By breaking down and rearranging the photographs' original message, the photomontage makes the viewer conscious of the arbitrariness of their normal message, making apparent ideological

[39] It is worth noting in relation to this that the Zürich group was undoubtedly influenced by the fact that Switzerland was neutral during the First World War. Zürich thus attracted many pacifists, or other people opposed to the war. For the members of the Berlin group, on the other hand, the terrors of war had been a much more direct experience. It is my guess that this made them less inclined to theoretically distance themselves to the degree of Tzara and his colleagues.

modes of deception. This effect would not have been as powerful if the objects or people were painted, for example, as they would then lack the indexicality pointing back to the real world outside of the work, which is a defining trait of photography. The strength of photomontage thus lies in its ability to *de-naturalize* things by turning upside-down the normal notions of realism in photography, creating new, more complex messages. As photography scholar Michel Frizot has brilliantly summarized this effect, "the only logic that photomontage obeys is that of referential substitution, in which meaning is cast aside by virtue of the lie inherent in the artefact" (1991).

In ImageNet Roulette, *How do you see me?* and *Spirit is a Bone*, it is the database – along with the practice of categorizing people as a whole – which is being exposed as a lie by making apparent the arbitrariness of grounding personal identity in the face or skull. By extension, this has some interesting consequences for the genre of the portrait more generally. Art historian Benjamin Buchloh describes in the essay *Residual Resemblance*, how the early 1900s constituted a time of turmoil for the artistic portrait as a consequence of how it had been disassembled by the Cubists:

> These antiportraits fuse the sitter's subjectivity in a continuous network of phenomenological interdependence between pictorial surface and virtual space, between bodily volume and painterly texture, as all physiognomic features merge instantly with their persistent negation in a pictorial erasure of efforts of mimetic resemblance (1994).

After Picasso completely de-naturalized the epistemic condition which saw mimetic resemblance as necessary for representing subjectivity, contesting physiognomic truth as a natural given, the genre was in crisis. While the Neue Sachlichkeit movement sought to salvage the idea of pictorial resemblance in art (August Sander can definitely be read into this tendency, as Buchloh indeed does), the Dadaist photomonteurs exploited this newfound confusion surrounding the photographic portrait. This can be seen in the prevalence of cut-out humans in the montages. My argument is that it was precisely the complicated status of the photographic portrait which allowed such playful use. On the one hand, a portrait could simply refer to *any* person and thus function to situate the broader message within the social realm in general. On the other hand, it was the inherent inclination of most viewers to seek to connect it to a specific identity which made the total composition of the photomontages all the more powerful.

The most famous of the Berlin photomonteurs was John Hartfield, who spent most of his career producing photomontages for Communist magazines and pamphlets. His oeuvre can thus be understood as political propaganda; a notion which he personally embraced. This is

evident in the manifesto authored by Hartfield and the other members of the so-called Red Group, where they state that "a good Communist is first of all a Communist, and only secondarily a technician, artist and so on" (Grosz, Witte, & Heartfield 2011 [1924]: 239). Thus, in line with Bolshevik doctrine, the majority of Heartfield's work was aimed against fascism, but also often against the Social Democrats. It should be noted, however, that in his explicit party affiliation, Heartfield stood out from the rest of the Berlin Dada group, of which only he and two others ever joined the German Communist party (KDP) (Ades 1986: 26). If we nevertheless take his photomontage *Adolf der Übermench: Schluckt Gold und Redet Blech* (1932) (Fig. 27) as an example of the above discussion, it offers a double reading: first, it can be understood as a critique of how Hitler – with his chest and belly full of gold – was motivated by the interests of his financial backers. But it can just as well be read more broadly as a critique of the NSDAP – or even of the contemporary political system as a whole.

The same could be said about *ImageNet Roulette* and *Spirit is a Bone.* While the photographic portraits allow the audience to perceive the databases as a collection of *individuals*, the serial presentation simultaneously allowed me to project *myself* into the depicted collective. The works thus transcend the idea of subjectivity as inherently linked to pictorial representation. This was achieved through a serial use of photography in concert with other curatorial choices. Along the same lines as Dadaist photomontage, the three works thus stand as a critique of AFR technology in general, and not just the specific systems which have been applied. Furthermore, through the act of re-contextualizing the datasets into the gallery space, the works serve as an institutional critique in line with the Duchampian readymade: in this respect, however, it is the institution of national surveillance rather than art itself which is the object of scrutiny.

What the three projects also share with Dadaism and Surrealism is that they specifically address the philosophical issues inherent positivist objectivity, which can be said to have peaked in terms of recognition during the early 1900s. As positivist disciplines were being debunked one after another, so are biometric technologies currently under fire as academics, journalists, politicians and the general public are becoming increasingly aware of the dangers they constitute. AFR, like the portrait genre in the 1910s, can thus be said to face an epistemic crisis almost exactly a century after the first wave of mimetic skepticism flushed Europe. This time around, however, Marxist dialectical materialism no longer seems a valid alternative as its Hegelian foundation has lost its previous status. The critique thus appears less as a clash between two opposing ideologies and more as a deconstruction of the one. Reviving the discussion from Chapter 1, I therefore think it suitable to describe the works as carving out the

foundations for a critical viewing ecology; not by offering an alternative to the present, but by scrutinizing the epistemic basis of present-day surveillance systems and engendering a critical attitude towards AFR. In this sense, they can be said to promote what has been termed a *new camera consciousness* characterized by heightened public awareness and literacy towards camara-based surveillance (McCosker & Wilken 2020). The title *Training Humans*, in this perspective, seems all the more appropriate.

In light of the above, it should be stressed that ImageNet Roulette, *How do you see me?* and *Spirit is a Bone* avoid solicitating the kind of iconoclastic response aimed for by (most) Dadaists, who were anti-everything. The projects also stand clear of the passifying effects of the mathematical sublime discussed in Chapter 1. Instead, I propose that the works invite what the artist James Bridle has described as a *Glomar Response*; a term which in US law describes a way of responding to requests for information which neither confirms nor denies the existence of the information in question. Facing matters of global surveillance, Bridle argues, a Glomar Response might be the only reaction that is true to the complexity of the world (Shipwright 2015). In this sense, the art projects invite dialectical reflection as an antidote for the positivism they critique. By poetically engaging with the matter at hand, the works of Paglen, Dewey-Hagborg and Broomberg & Chanarin let the audience fill in the blanks themselves. The message, if not the aesthetic expression per se, is thus akin to that of Surrealism, which had free association as one of its founding principles. Importantly, they achieve this without receding to the ironic distance typical of the early twentieth century anti-artists.

Before rounding off, I will comment on what might otherwise appear as a paradox. It is clear that the works share with Dadaism a concern with contesting ideas originating in the Enlightenment project. In light of this, one might ask – as Melissa Gronlund has done – whether there is an inert problem in this endeavor, as the theoretical construction of aesthetics is *itself* an Enlightenment idea. In a broad sense, the epistemic condition being critiqued is essentially the same one that made way for art's privileged position as a mode of human expression. Reflecting on this, Gronlund asks: "if art is an Enlightenment object, the sustained challenge of current philosophies to the Enlightenment imperils the category of art" (2017: 99). This, of course, holds true for the works discussed in this thesis as well. Engaging with them, however, this problem was never experienced by me as contradictory. Yes, the works were made and presented within an art-specific context, but not at one point did their claim to being art occur to me as relevant – in spite of their radically "un-arty" appearance. On the contrary, the works' status as *art objects* appeared subordinate to their status as *epistemic objects*; a conception enabled by Dadaist legacy. The fact that this supposed opposition never appeared as a conflict

– despite this being a hot topic within contemporary art – is precisely what makes ImageNet Roulette, *How do you see me?* and *Spirit is a Bone* so good art. *This* is the real paradox.

Figure 27: *John Heartfield,* Adolf der Übermensch: Schluckt Gold und redet Blech *(1932). Photomontage. Retrieved on 22.05.2021 from Johnheartfield.com. URL: https://www.johnheartfield.com/John-Heartfield-Exhibition/john-heartfield-art/famous-anti-fascist-art/heartfield-posters-aiz/adolf-the-superman-hitler-portrait.*

Conclusion

Divided in two parts, this chapter has first discussed AFR from a historico-political perspective before proceeding to an analysis of the aesthetic methods used to engender these reflections. It is clear from this discussion that current AFR technology is the latest stage in an evolution which has its starting point in the physiognomic disciplines of the 1800s. The widespread belief in AFR technologies' claim to truth can be partially explained by sheer gradual naturalization, and partially by the prevalence of Kantian idealism in Western thought. It is evident that image

recognition, like Descartes and Kant, treat form as separable from content. When the forms are faces, this practice evokes questions of identity and its relation to the body. That an axiomatic system of truth production should appear as particularly suited to a discipline based on mathematical calculations is hardly surprising. Neither is this a problem in itself, but serious problems nevertheless arise when the calculations are confused as objective representations of reality. The art projects presented in this thesis serve to mystify the truth claims of AFR as well as the prevalence of identificatory schemes more generally. I have argued that the development of optical technologies as a means of control can be traced back to a period characterized by increased mobility – also within the visual field – and that these tendencies have been amplified by the War on Terror.

AFR systems are more or less intentionally designed to reify the systems of power they are made to serve: the attributed meanings – even if completely unfounded – can become true in the sense that they shape how people are consequently treated by various institutions, thus reinforcing existing power structures. In light of the deterministic effects of categorization in AFR, claims of objectivity are especially dangerous. This is why it is so important to de-naturalize AFRs role in society, which, as I have argued, has been achieved in ImageNet Roulette, *How do you see me?* and *Spirit is a Bone* by means reminiscent of Dadaist photomontage. The effect is that the works discussed in this thesis offer an experience of disruption of the sense of flow associated with digital phenomena in general, and photography in particular. In this way, they enable a critical "look back" at a kind of technology which is increasingly being ingrained in most areas of modern life.

In conclusion, I propose for the reader to engage in a little thought experiment. In a special issue of the journal AI & Society themed *ways of machine seeing*, the editors pose the following question: "can machine vision step beyond the 'ocularcentric' metaphysics of the Western gaze and the reproduction of racial capital?" (Azar, Cox, & Impett 2021). While this is certainly an interesting topic for discussion, I am skeptical towards the fruitfulness of this approach. Knowing that the practical and economic basis for producing AFR has always been to perpetuate existing power structures, I round off by rephrasing the question: imagine a non-Western (or westernized), non-ocularcentric society without race-, class- and gender inequality. What incentives would exist in such a scenario to develop automated facial recognition technology in the first place?

Concluding Remarks

I introduced this thesis with a story about my visit at the Pompidou Centre and how my smart-phone's camera app detected painted faces in addition to real ones. Since then, I have bought myself a new phone. Curious as to whether its much-updated software would do the same as the previous one, I went to the Edvard Munch collection at the KODE Museum in Bergen and reenacted the experiment from Paris. Once again holding the camera up in front of me, it soon became clear that just like before, my new phone indicated faces with yellow squares – even in some very blurred pictures, such as *At the Deathbed* (1895) where one woman's face is depicted simply as a white oval with two black dots for eyes (Fig. 28). At the same time, it indicated none in another painting where almost all of the faces share a similar look (Fig. 29). This serves to underscore the continued opacity of the logic informing AFR as it presents itself to the human beholder.

When it comes to *recognition* of faces, however, much has happened between my old model from 2012 and my new iPhone 11 Pro Max. Not only have facial recognition features become normalized only in the last few years, but this particular model is known for being amongst the most reliable on the market in relation to its facial detection feature's accuracy. I was therefore intrigued to hear that a fellow student's son had managed to open hers (she has the same model) by presenting it with a photograph of his mother. I have been unsuccessful, however, in reenacting this stunt, although not for a lack of trying. As various online commentators argue, such a trick *should* be unable to fool the newest iPhones, as the system is programmed to perceive the difference between two-dimensional and three-dimensional shapes. Exactly how it does so is unknown to the public, however, in being a trade secret. In light of this, I find it especially curious what happened to me one day. Studying myself in the bathroom mirror, an idea occurred to me: what if my mirror-image is detailed enough to fool the phone even though it is technically as flat as a photograph? I raised my hand with the locked phone and turned the selfie-camera towards my reflection while simultaneously covering the other camera turned against my real face. To my great surprise, it was unlocked.

Reflecting on the above, the famous remark once made by Pablo Picasso comes to mind: "computers are useless. They can only give you answers." Having analyzed computational processes for two years now, it seems that even though they can indeed give some answers, they give rise to just as many questions. As an art historian, I also cannot help but question how the findings in this thesis reflect on the legacy of the artistic avant-garde more generally. As

Jonathan Crary has noted in his much-cited book *Techniques of the Observer*, the lessons from the fabled break with optical realism taking place in the mid-1800s were probably restricted to a much smaller circle of especially interested people than what traditional art historical narratives suggest. He argues that contrary to an alleged "perceptual revolution" informing so many aspects of the history of Western art, the innovations in modernist painting were rather occurring on the *margins* of society. Simultaneously as these developments were taking place, Crary points out, a "vast hegemonic organization of the visual" became increasingly powerful throughout the twentieth century "with the diffusion and proliferation of photography, film, and television" (1991: 4). The current reliance on photography as a document of truth in AFR systems serves to confirm that this did not stop in the twentieth century.

While Crary's argument precedes mine by thirty years, I nevertheless think that the discussion throughout this thesis has added new dimensions to his skepticism towards the cultural effects of modernist art. While this particular issue has not been a conscious topic of this thesis, the consequences are nevertheless too important to go unnoted. As I have made clear by now, ideas concerning optical realism and the documentary value of photography still hold sway within the computer sciences. In conclusion then, I ask whether letting automated systems of image analysis continue to inform public and private decisions, contributes to the continued proliferation of an epistemic system which enables only limited and biased understandings of the world – especially concerning the study of humans. Rounding off, I once again underscore the importance of countering these tendencies with increased artistic and scholarly scrutiny. While leaving still many questions to be answered, I hope that this thesis has inspired the reader at least a little to contribute to this endeavor in the future.

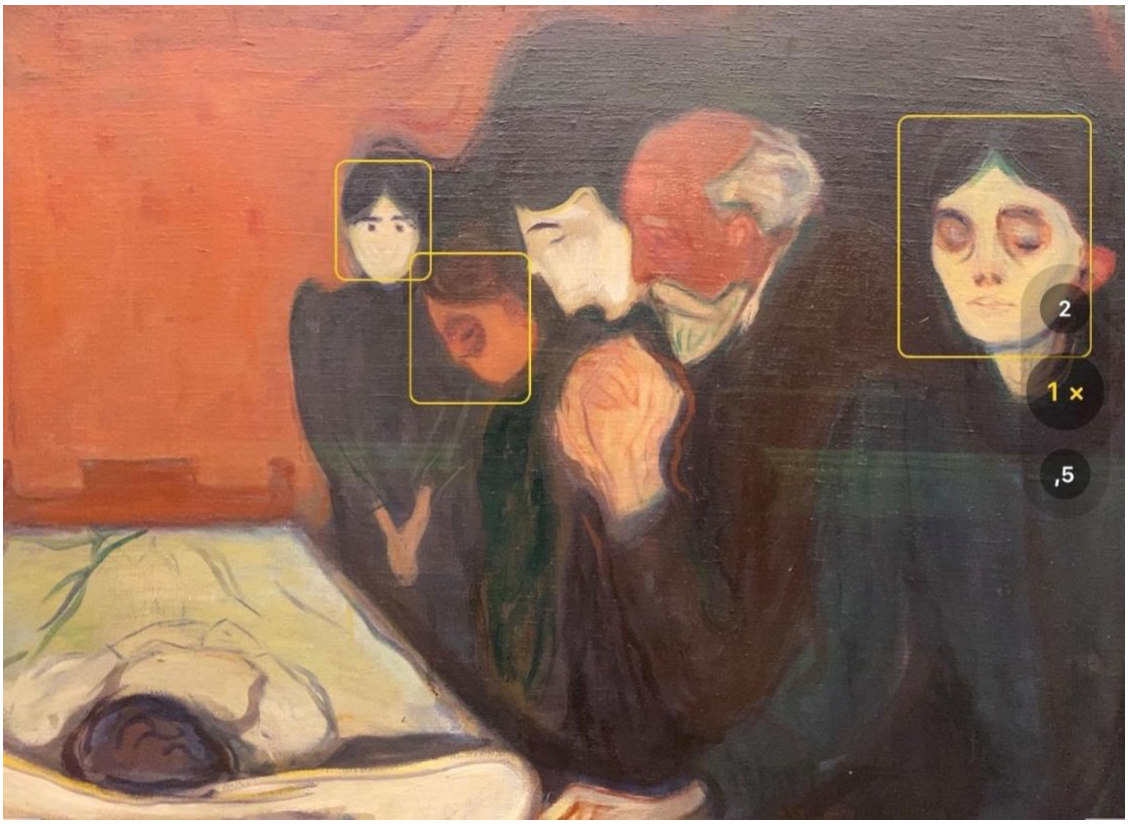


Figure 28: *Detail of* At the Deathbed *(1895) by Edvard Munch. Screenshot of my phone's camera app.*

Figure 29: *Detail of* Evening on Karl Johan *(1892) by Edvard Munch. Screenshot of my phone's camera app.*

Bibliography

Ades, Dawn (1986). *Photomontage*. 2. ed. London: Thames & Hudson.

Almås, Gry Blekastad (05.02.2019). "Digitalt diktatur: Kina planlegger sosialt poengsystem." *NRK*. Accessed 13.04.2019. URL: https://www.nrk.no/urix/kinas-digitale-diktatur_-gar-du-pa-rodt-lys_-blir-du-uthengt-pa-storskjerm-1.14369439.

Artforum (24.09.2019). "Image Database Purges 600K Photos After Trevor Paglen Project Reveals Biases ". *Artforum*. Accessed 17.01.2021. URL: https://www.artforum.com/news/image-database-purges-600k-photos-after-trevor-paglen-project-reveals-biases-80829.

Azar, Mitra, Geoff Cox, and Leonardo Impett (2021). "Introduction: ways of machine seeing." *AI & Society*. DOI: https://doi.org/10.1007/s00146-020-01124-6.

Balko, Radley (10.06.2020). "There's overwhelming evidence that the criminal justice system is racist. Here's the proof." *The Washington Post*. Accessed 24.04.2021. URL: https://www.washingtonpost.com/graphics/2020/opinions/systemic-racism-police-evidence-criminal-justice-system/.

BBC (24.06.2020). "Facial recognition to 'predict criminals' sparks row over AI bias." *BBC Technology*. Accessed 24.04.2021. URL: https://www.bbc.co.uk/news/technology-53165286.

Benjamin, Walter (2008 [1936]). *The Work of Art in the Age of Mechanical Reproduction*. Translated by J. A. Underwood. London: Penguin Books.

Berger, John (2013 [1969]). "Political Uses of Photomontage." In *Understanding a Photograph*, edited by Geoff Dyer: 28-32. New York, N.Y.: Aperture Foundation.

Bischoff, Paul (17.05.2021). "Surveillance camera statistics: which cities have the most CCTV cameras?". *Comparitech*. Accessed 23.05.20201. URL: https://www.comparitech.com/blog/vpn-privacy/the-worlds-most-surveilled-cities/.

Brilliant, Richard (1991). *Portraiture*. London: Reaktion Books.

Broca, Paul (1861). "Sur le volume et la forme du cerveau suivant les individus et suivant les races." *Bulletin Société d'Anthropologie Paris* 2:139-207. Cited in Gould, Stephen Jay (1997). *The Mismeasure of Man*. London: Penguin Books.

Broomberg, Adam, and Oliver Chanarin (2015). *Spirit is a Bone*. London: Mack.

Broussard, Meredith (2018). *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge, Massachusetts: MIT Press.

Buchloh, Benjamin H. D. (1994). "Residual Resemblance: Three Notes on the End of Portraiture." In *Face-Off: The Portrait in Recent Art*, edited by Melissa E. Feldman: 53-69. Philadelphia: Institute of Contemporary Art, University of Pennsylvania.

Buck, Louisa "After more than two decades together, artist duo Broomberg and Chanarin commit 'creative suicide'." *The Art Newspaper*. Accessed 24.03.2021. URL: https://www.theartnewspaper.com/news/after-more-than-two-decades-of-making-art-together-broomberg-and-chanarin-commit-career-suicide.

Byrne, Patrick, and Karen Hudson-Edwards (28.08.2018). "Three ways making a smartphone can harm the environment." *The Conversation*. Accessed 16.05.2021. URL: https://theconversation.com/three-ways-making-a-smartphone-can-harm-the-environment-102148.

Clarke, Graham (1992). "Public faces, Private Lives: August Sander." In *The Portrait in Photography*, edited by Graham Clarke: 71-93. Seattle: Reaktion Books.

Crary, Jonathan (1991). *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*. Cambridge, MA & London: MIT Press.

Crawford, Kate (30.05.2014). "The Anxieties of Big Data." *The New Inquiry*. Accessed 04.05.2021. URL: https://thenewinquiry.com/the-anxieties-of-big-data/.

_____ (2016). "Asking the Oracle." In *Astro Noise: A Survival Guide for Living under Total Surveillance*, edited by Laura Poitras: 138-51. New Haven & London: Whitney Museum of American Art & Yale University Press.

Crawford, Kate, and Trevor Paglen (n.d.). "Excavating AI: The Politics of Images in Machine Learning Training Sets ". Accessed 17.01.2021. URL: https://excavating.ai/.

Dagbladet (31.03.2020). "Dette hastverket er livsfarlig." *Dagbladet*. Accessed 13.04.2021. URL: https://www.dagbladet.no/kultur/dette-hastverket-er-livsfarlig/72313681.

Derrida, Jacques (1996). *Archive Fever: A Freudian Impression*. Translated by Eric Prenowitz. Chicago & London: The University of Chicago Press.

Dewey-Hagborg, Heather (2019). "How do you see me?", Accessed 28.02.2021. URL: http://deweyhagborg.com/projects/how-do-you-see-me

_____ (n.d.). "Bio." DeweyHagborg.com. Accessed 28.02.2021

Digicult (12.09.2019). "“Training Humans”: an exhibition by Kate Crawford and Trevor Paglen." *Digicult*. Accessed 28.05.2021. URL: http://digicult.it/slider/training-humans-an-exhibition-by-kate-crawford-and-trevor-paglen/.

Dreyfus, Hubert L. (1992). *What Computers Still Can’t Do*. Cambridge, MA: The MIT Press.

Döblin, Alfred (1983 [1929]). "Von Geschichtern Bildern und ihrer Warheit." In *Antlitz der Zeit: Sechzig Aufnahmen Deuthscher Menschen Des 20. Jarhunderts*, edited by August Sander. Munich: Schirmel/ Mosel.

Edwards, Elizabeth (2011). "Tracing Photography." In *Made to Be Seen: Perspectives on the History of Visual Anthropology*, edited by Marcus Banks and Jay Ruby: 159-89. Chicago and London: The University of Chicago Press.

Farocki, Harun (2004). "Phantom Images." *Public* No. 29:12-22.

Fondazione Prada (2019). "Kate Crawford | Trevor Paglen: Training Humans." Fondazione Prada. Accessed 28.02.2021. URL: http://www.fondazioneprada.org/project/training-humans/?lang=en

Foucault, Michel (1999 [1975]). *Overvåkning og Straff: Det Moderne Fengsels Historie*. Translated by Dag Østerberg. 3. ed. Oslo: Gyldendal.

_____ (2006 [1966]). *Tingenes Orden: En Arkeologisk Undersøkelse Av Vitenskapene Om Mennesket*. Oslo: Spartacus.

Frizot, Michael (1991). "The Lying Camera." In *Photomontage*, edited by Michael Frizot. London: Thames & Hudson Ltd.

Fuller, Matthew (2014). "Data." In *Johns Hopkins Guide to Digital Media*, edited by Marie-Laure Ryan, Lori Emerson and Benjamin J. Robertson: 125-27. Baltimore: Johns Hopkins University Press.

Gates, Kelly A. (2011). *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*. New York & London: New York University Press.

Gibson, William (1984). *Neuromancer*. New York: Ace.

Ginzburg, Carlo, and Anna Davin (1980). "Morelli, Freud and Sherlock Holmes: Clues and Scientific Method." *History Workshop* No. 9:5-36.

Goggin, Benjamin (16.01.2018). "We Need a Bigger Palette: Is Google's Arts And Culture App Racist?". *Digg*. Accessed 27.02.2021. URL: https://digg.com/2018/google-arts-culture-racist-face.

Gould, Stephen Jay (1997). *The Mismeasure of Man*. London: Penguin Books.

Gronlund, Melissa (2017). *Contemporary Art and Digital Culture*. New York, NY: Routledge.

Grosz, George, Karl Witte, and John Heartfield (2011 [1924]). "Manifesto of the Communist Artists' Group." In *100 Artists' Manifestos: From the Futurists to the Stuckists*, edited by Alex Danchev: 239-40. London: Penguin Books.

Haggerty, Kevin D., and Richard V. Ericson (2000). "The Surveillant Assemblage." *British Journal of Sociology* 51, No. 4:605-22.

Hegel, G. W. F. (2018 [1807]). *The Phenomenology of Spirit*. Translated by Michael Inwood. Oxford: Oxford University Press.

Hegel, Georg W. F. (2009 [1807]). *Åndens fenomenologi*. Oslo: Pax Forlag A/S.

Heim, Michael (2014). "Virtuality." In *The Johns Hopkins Guide to Digital Media*, edited by Marie-Laure Ryan, Lori Emerson and Benjamin J. Robertson: 514-19. Baltimore: Johns Hopkins University Press.

Henning, Michelle (2018). *Photography: The Unfettered Image*. London & New York: Routledge.

Hill, Kashmir (24.01.2020). "Wrongfully Accused by an Algorithm." *The New York Times*. Accessed 24.04.2021. URL: https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

Holert, Tom (2016). "Exposure without Intelligibility." In *Watched! Surveillence, Art and Photography*, edited by Ann-Christin Bertrand, Beate Cegielska, Jacob Fabricius, Dragana Vujanovic, Louise Wolthers and Niclas Östlind: 276-81. Cologne, Germany: Verlag Der Buchhandlung Walther Konig

Hopkins, David (2018). *After Modern Art: 1945-2017*. Oxford: Oxford University Press; 2nd edition.

Hutson, Matthew (03.05.2018). "AI researchers allege that machine learning is alchemy." *Science Magazine*. URL: https://www.sciencemag.org/news/2018/05/ai-researchers-allege-machine-learning-alchemy.

Jay, Martin (1988). "Scopic Regimes of Modernity." In *Vision and Visuality*, edited by Hal Foster: 3-23. Seattle: Bay Press.

_____ (1994). *Downcast Eyes: The Denigration of Vision in Twentieth-Century French Thought*. Berkeley, Los Angeles & London: University of California Press.

Jenkins, Richard (2012). "Identity, surveillance and modernity: Sorting out who's who." In *Routledge Handbook of Surveillance Studies*, edited by Kirstie Ball, Kevin D. Haggerty and David Lyon: 159-66. Abingdon, Oxon: Routledge.

Jurgenson, Nathan (2011). "Digital Dualism versus Augmented Reality." *Cyborgology*, 14.05.2021. URL: https://thesocietypages.org/cyborgology/2011/02/24/digital-dualism-versus-augmented-reality/.

Karn, Ujjwal (2016a). "An Intuitive Explanation of Convolutional Neural Networks." 11.08.2016. URL: https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/.

_____ (2016b). "A Quick Introduction to Neural Networks." 09.08.2016. URL: https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/.

La Fors-Owczynik, Karolina, and Irma van der Ploeg (2016). "Migrants at/as Risk: Identity Verification and Risk.Assessment Technologies in the Netherlands." In *Digitizing Identities: Doing Identity in a Networked World*, edited by Irma van der Ploeg and Jason Pridmore. New York, NY: Routledge.

Larsen, Jonas, and Mette Sandbye (2014). "The New Face of Snapshot Photography." In *Digital Snaps: The New Face of Photography*, edited by Peter Larsen and Mette Sandbye: xv-xxxii. London: Tauris.

Latour, Bruno, Pablo Jensen, Tommaso Venturini, Sébastian Grauwin, and Dominique Boullier (2012). "'The whole is always smaller than its parts' – a digital test of Gabriel Tardes' monads." *The British Journal of Sociology* Vol. 63, No. 4:590-615.

Lee-Morrison, Lila (2019). *Portraits of Automated Facial Recognition: On Machinic Ways of Seeing the Face*. Bielefeld: Transcript Verlag.

Lil Uzi Hurt (18.09.2019). "No matter what kind of image I upload, ImageNet Roulette, which categorizes people based on an AI that knows 2500 tags, only sees me as Black, Black African, Negroid or Negro. Some of the other possible tags, for example, are “Doctor,” “Parent” or “Handsome”." Twitter, 18.09.2019, URL: https://twitter.com/lostblackboy/status/1174112872638689281/photo/2

Lister, Martin (2014). "Overlooking, Rarely Looking and Not Looking." In *Digital Snaps: The New Face of Photography*, edited by Jonas Larsen and Mette Sandbye: 1-23. London: Tauris.

Lyon, David (2002a). "Introduction." In *Surveillance As Social Sorting: Privacy, Risk and Automated Discrimination*, edited by David Lyon. London & New York: Routledge.

_____ (2002b). "Surveillance as social sorting: computer codes and mobile bodies." In *Surveillance As Social Sorting: Privacy, Risk and Automated Discrimination*, edited by David Lyon: 13-30. London & New York: Routledge.

Magnet, Shoshana A. (2018). "When Biometrics Fail: Gender, Race, and the Technology of Identity." In *Surveillance Studies: A Reader*, edited by Torin Monahan and David Murakami Wood: 116-19. New York: Oxford University Press.

Mallonee, Laura "Creepy Portraits Made With Even Creepier Surveillance Tech." *Wired Magazine*, 16.02.2016.Accessed Date Accessed. URL: https://www.wired.com/2016/02/adam-broomberg-oliver-chanarin-spirit-is-a-bone/.

McCosker, Anthony, and Rowan Wilken (2020). *Automating Vision: The Social Impact of the New Camera Consciousness*. New York & London: Routledge.

Metz, Cade (20.09.2019). "‘Nerd,’ ‘Nonsmoker,’ ‘Wrongdoer’: How Might A.I. Label You?". *The New York Times*. Accessed 27.02.2021. URL: https://www.nytimes.com/2019/09/20/arts/design/imagenet-trevor-paglen-ai-facial-recognition.html

Mirlesse, Sabine (2014). "Adam Broomberg and Oliver Chanarin." *BOMB* No. 130:32-40.

Monahan, Torin, and David Murakami Wood (2018a). "Introduction to Section 4: Identity and Identification." In *Surveillance Studies: A Reader*, edited by Torin Monahan and David Murakami Wood: 93-95. New York, NY: Oxford University Press.

_____ (2018b). "Introduction to Section 5: Borders and Mobilities." In *Surveillance Studies: A Reader*, edited by Torin Monahan and David Murakami Wood: 121-23. New York, NY: Oxford University Press.

Morris-Reich, Amos (2016). *Race and Photography: Racial Photography as Scientific Evidence, 1876-1980*. Chicago and London: The University of Chicago Press.

Mortensen, Mette (2012). *Kampen om ansigtet: fotografi og identifikasjon*. Copenhagen: Museum Tusculanums Forlag.

Mousse Magazine (n.d.). "Kate Crawford and Trevor Paglen “Training Humans” at Osservatorio Fondazione Prada, Milan." *Mousse Magazine*. Accessed 28.02.2021.

URL: http://moussemagazine.it/kate-crawford-and-trevor-paglen-training-humans-at-osservatorio-fondazione-prada-milan-2019-2020/.

Nelson, Maggie (2015). *The Argonauts*. Minneapolis: Graywolf Press.

Nilsen, Amund Aune, and Håkon Olaussen (30.01.2012). "Å inngå et terrorforbund er å gjøre mer enn å bare tenke tanken." *NRK*. Accessed 12.05.2021. URL: https://www.nrk.no/norge/historisk-dom-om-terrorplanlegging-1.7974329.

Nowviskie, Bethany (2014). "Algorithm." In *The Johns Hopkins Guide to Digital Media*, edited by Marie-Laure Ryan, Lori Emerson and Benjamin J. Robertson: 1-4. Baltimore: Johns Hopkins University Press.

O'Hagan, Sean (05.02.2016). "Negative humanity: the birth of the digital death mask." *The Guardian*. Accessed 28.05.2021. URL: https://www.theguardian.com/artanddesign/2016/feb/05/digital-death-mask-photography-broomberg-and-chanarin-spirit-is-a-bone.

Obrist, Hans Ulrich (01.02.2016). "What Is the Future of Art?". *Artsy*. URL: https://www.artsy.net/article/hans-ulrich-obrist-the-future-of-art-according-to-hans-ulrich-obrist.

Pardo, Alona (2019). "From 'Car' to 'Omelette' - Trevor Paglen in Conversation with Alona Pardo." In *Trevor Paglen: From Apple to Anomaly*, edited by Sarah Cook and Alona Pardo: 35-45. London, United Kingdom: Barbican.

Paul, Christiane (2015). *Digital Art*. 3. ed. London: Thames & Hudson.

Popper, Frank (1993). *Art of the Electronic Age*. London: Thames and Hudson.

Rahimi, Ali "*Ali Rahimi's talk at NIPS (NIPS 2017 Test-of-time award presentation)"*. Video documentation of talk. URL: https://www.youtube.com/watch?v=Qi1Yry33TQE.

Rancière, Jacques (2012 [2008]). *Den emansiperte tilskuer*. Translated by Geir Uvsløkk, *Pax Artes*. Oslo: Pax Forlag.

Rea, Naomi (23.09.2019). "How ImageNet Roulette, an Art Project That Went Viral by Exposing Facial Recognition's Biases, Is Changing People's Minds About AI." *Artnet*. Accessed 27.02.2021. URL: https://news.artnet.com/art-world/imagenet-roulette-trevor-paglen-kate-crawford-1658305.

Rees, Tobias (2020). "Machine/Intellligence: On the Philosophical Stakes of AI Today." In *Beyond the Uncanny Valley: Being Human in the Age of AI*, edited by Claudia Schmuckli, Yuk Hui and Janna Keegan: 110-17. California: Cameron & Company Inc.

Rubinstein, Daniel, and Katarina Sluis (2013). "The digital image in photographic culture: algorithmic photography and the crisis of representation." In *The Photographic Image in Digital Culture,*, edited by Martin Lister: 22-40. London & New York: Routledge.

Russell, Legacy (2020). *Glitch Feminism: A Manifesto*. London: Verso.

Ryan, Marie-Laure (2014). "Cyberspace." In *The Johns Hopkins Guide to Digital Media*, edited by Marie-Laure Ryan, Lori Emerson and Benjamin J. Robertson: 118-20. Baltimore: Johns Hopkins University Press.

Rzepecky, Sophie (September 2019). "Training Humans." *DAMN° Magazine*. Accessed 28.02.2021. URL: https://www.damnmagazine.net/2019/09/25/training-humans-fondazione-prada/.

Salter, Mark (2018). "Passports, Mobility, and Security: How Smart Can the Border Be?" In *Surveillance Studies: A Reader*, edited by Torin Monahan and David Murakami Wood: 127-32. New York, NY: Oxford University Press.

Sanooei, Pouya (2017). "What is the difference between deep learning and usual machine learning?" GitHub. [Thread post.], Last Modified 14.10.2017, Accessed 28.02.2021. URL: https://github.com/rasbt/python-machine-learning-book/blob/master/faq/difference-deep-and-normal-learning.md.

Sekula, Alan (1981). "The Traffic in Photographs." *Art Journal* Vol. 41, No. 1:15-21.

_____ (1986). "The Body and the Archive." *October* Vol. 39:3-64.

Shipwright, Fiona (30.07.2015). "Data Ontology: James Bridle on secrecy, surveillance, and the limits of knowledge." *Rhizome*. Accessed 12.05.2021. URL: https://rhizome.org/editorial/2015/jul/30/interview-james-bridle/.

Simonite, Tom (11.01.2018). "When It Comes to Gorillas, Google Photos Remains Blind." *Wired*. Accessed 27.02.2021. URL: https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/?redirectURL=https%3A%2F%2Fwww.wired.com%2Fstory%2Fwhen-it-comes-to-gorillas-google-photos-remains-blind%2F.

Snowden, Edward (2016). "Astro Noise." In *Astro Noise: A Survival Guide for Living under Total Surveillance*, edited by Laura Poitras: 120-25. New Haven & London: Whitney Museum of American Art & Yale University Press.

Sontag, Susan (2010 [1977]). *On Photography*. London: Penguin Books.

Stalder, Felix, and David Lyon (2002). "Electronic identity cards and social classification." In *Surveillance As Social Sorting : Privacy, Risk and Automated Discrimination*, edited by David Lyon: 77-93. London & New York: Routledge.

Steyerl, Hito (2019). "A Sea of Data: Pattern Recognition and Corporate Animism (Forked Version)." In *Pattern Discrimination*, edited by Clemens Apprich, Wendy Hiu Kyong Chun, Florian Cramer and Hito Steyerl: 1-22. Minneapolis & London: University of Minnesota Press.

Tadviser (n.d.). "Vocord FaceControl 3D." URL: http://tadviser.com/index.php/Product:Vocord_FaceControl_3D.

Tagg, John (1988). *The Burden of Representation: Essays on Photographies and Histories*. London: Macmillan.

The Photographers' Gallery (26.09.2019). "*How Do You See Me? An Interview with Heather Dewey-Hagborg"*. Video. URL: https://www.youtube.com/watch?v=_68u_uEY4FM&t=217s&ab_channel=ThePhotographers%27Gallery.

_____ (29.11.2019). "*What is the genealogy of Training Humans? An interview with Kate Crawford & Trevor Paglen"*. Video. URL: https://www.youtube.com/watch?v=V02orRSAqVI.

Trottier, Daniel (2016). "Caring For the Virtual Self on Social Media: Managing Visibility on Facebook." In *Digitizing Identities*, edited by Irma van der Ploeg and Jason Pridmore: 60-76. New York, NY: Routledge.

Turing, Alan M. (1937 [1936]). "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* Vol. 2-42, No. 1:230-65.

_____ (1950). "Computing Machinery and Intelligence." *Mind* Vol. 59, No. 236:433-60.

Tzara, Tristan (2011 [1918]). "Dada Manifesto." In *100 Artists' Manifestos: From the Futurists to the Stuckists*, edited by Alex Danchev: 136-44. London: Penguin Books.

Tørresen, Jim (2013). *Hva er kunstig intelligens.* Oslo: Universitetsforlaget.

VocordCompany (16.09.2016). "*3D facial recognition system VOCORD FaceControl 3D"*. URL: https://www.youtube.com/watch?v=FRq0qXOoBAc.

Weller, Toni (2012). "The information state: An historical perspective on surveillance." In *Routledge Handbook of Surveillance Studies*, edited by Kristie Ball, Kevin D. Haggerty and David Lyon: 57-63. Oxon: Routledge.

West, Shearer (2004). *Portraiture.* Oxford & New York: Oxford University Press.

Wolthers, Louise (2016). "Viewing Ecologies." In *Watched! Surveillence, Art and Photography*, edited by Ann-Christin Bertrand, Beate Cegielska, Jacob Fabricius, Dragana Vujanovic, Louise Wolthers and Niclas Östlind: 282-89. Cologne: Verlag Der Buchhandlung Walther Konig

Wong, Julia Carrie (18.09.2019). "The viral selfie app ImageNet Roulette seemed fun – until it called me a racist slur." *The Guardian*. Accessed 27.02.2021. URL: https://www.theguardian.com/technology/2019/sep/17/imagenet-roulette-asian-racist-slur-selfie

Yang, Kaiyu, Klint Qinami, Fei-Fei Li, Jia Deng, and Olga Russakovsky (17.09.2019). "Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy." *ImageNet.org*. Accessed 17.01.2021. URL: http://image-net.org/update-sep-17-2019.

Zhou, Zhenglong, and Chaz Firestone (2019). "Humans can decipher adversarial images." *Nature Communications* Vol. 10.

Žižek, Slavoj (2008 [1989]). *The Sublime Object of Ideology*. London: Verso.

Østerberg, Dag (2009). "Innledning til *Åndens Fenomenologi*." In *Åndens fenomenologi*: 7-27. Oslo: Pax.

Aas, Katja Franko (2018). "'Crimmigrant' Bodies and Bona Fide Travelers: Surveillance, Citizenship, and Global Governance." In *Surveillance Studies: A Reader*, edited by Torin Monahan and David Murakami Wood: 138-42. New York, NY: Oxford University Press.