

Truth, Lies and ‘Deepfakes’: The Epistemology of Photographic Depictions

by

Amanda C. Prusila

A thesis submitted to the Faculty of Graduate Studies and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Master of Arts

in

Film Studies

Carleton University
Ottawa, Ontario, Canada

© 2022
Amanda C. Prusila

Abstract

When provided a large variety of photographs and videos, AI algorithms can generate synthetic, manipulable video representations of human faces to a significant degree of perceptual realism. These visual media, commonly referred to as ‘deepfakes,’ while often used for comical or educational purposes, are largely seen as serious threats to our understanding of truth. There is great concern that this technology will call into question the validity of all images, leading to an undermining of our trust in institutions and in the democratic process.

Though a modern technology, the fundamental questions ‘deepfakes’ raise are part of longstanding epistemological debates, namely: how to understand the relationship between medium, technology and message; and how to define and use truth and knowledge.

Within the context of such philosophical debates, I argue that these aforementioned negative consequences are all possible using ‘deepfakes’; however, they are not inherently deceptive or malicious. It is not the medium itself that poses these threats, but rather how humans use them. ‘Deepfakes’ do not show a radical departure in our conceptions of truth in representations, but they do demonstrate an overconfidence in the truth of photographic images. Rather than outright bans or other technology-specific regulation, I propose that the careful study of ‘deepfakes’ is an opportunity to address some lapses in critical accounts of imagery and visual representation more generally. Additionally, I urge us to move away from using such a loaded, inaccurate term as ‘deepfakes’ to refer to this technology.

Acknowledgements

Thank you to the faculty and staff at the School for Studies in Art and Culture. I never imagined academia could be so invigorating and inviting, and it has all been thanks to your passion and compassion.

Above all else, I owe a gargantuan thank you to my thesis supervisor, Dr. Marc Furstenau. Thank you for your unwavering support, for never once complaining about our meetings (and my time completing this degree) going well over time, and for your constant reassurance that, yes, this will eventually come to fruition with enough diligence and pragmatism. I could not have asked for a better mentor. Thank you.

Thank you to my cohort—Azarin, Eleanor, Gigi, Niusha and especially Rachel—for the comradery and wonderful academic insight. I would also like to thank my upperclassmen, namely Ethan, Kyle, Matt, Myles, Onoshe, Sneha and Yiwen, for their guidance during my first year.

Thank you to my family, Mike, Julie and Erkki, for showing me *Star Wars* and *Bicentennial Man* at a young age. These movies' depictions of AI captivated and terrified me, respectively, and no doubt led to my endless interest in the powers of technology and what it means to be human. Thank you, more importantly, for your unconditional love and support, especially throughout the thesis process, when my workload and a global health crisis kept us apart.

Max, thank you for your love, your patience, and for always believing in me. You make me a better person, and life is most fun with you. Thank you, Alfie, for being the sweetest old mutt.

My friends have been invaluable throughout this process. Thank you, Caitlin, for your wonderful sense of humour and grad-school guidance. Thank you to Jen, Alanna, Owen, Jack, Clarissa and Steve. Our great conversations and movie nights have kept me grounded when school, work and coronavirus angst were all-consuming. Thanks for never telling me to stop sending so many memes.

Thank you, AJ and Jillian, for treating me like a human first and an employee second.

It never escapes my mind how privileged I am to have access to education, which is too often out of reach for people in even the most prosperous countries. To quote Toni Morrison, “just remember that your real job is that if you are free, you need to free somebody else. If you have some power, then your job is to empower somebody else.” I want to always keep that in mind.

This thesis is dedicated to the late Sheila Prusila, for showing me the joy of lifelong education.

Table of Contents

Abstract.....	i
Acknowledgements	ii
List of Images	v
Introduction.....	1
“The Search for Truth and Understanding”.....	3
Unmasking the Truth.....	6
Same Spiel, Different Medium	8
The Rise of ‘Deepfakes’	11
My Argument.....	12
Chapter 1: Separating Truth from Technology and the Medium.....	17
Technology	26
Focus on Photography, Not Photographs.....	30
Primary Functions and Markings.....	31
Depiction	35
Detection	37
Automation and Intentionality.....	41
Medium.....	42
Materials	44
Creation	44
Experience	45
Essentialist Accounts of Photography	47
Summary	49
Chapter 2: Characterizing Truth as a Human-Made Action	51
Understanding What is (and is Not) a Type of Truth	52
Truth as Objective Reality	53
Truth as Subjective Experience	54
Optical Illusions.....	55
Perceptual Realism	56
Fiction and Non-fiction.....	57
Using Truth Claims	59
Other People’s Perceptions	59
Evaluating Truth Claims.....	61
Falsities versus Lies.....	67
Context.....	69
‘Deepfakes’ and “Cheap Fakes”	71
Summary	75
Chapter 3: ‘Decoding’ the Algorithm and New Media	76
Brief History of Computer-Generated Faces	78
Motion Capture.....	78
Performance Capture and Facial Mapping	79
Structuralism and Digital Media.....	82

The Analog and the Digital.....	83
Human Agency and Algorithms.....	85
Summary	87
Conclusion	89
Not all Doom and Gloom.....	89
Where to Go from Here	91
Bibliography	93

List of Images

Fig. 1. Francesca Panetta and Halsey Burgund. Still from *In Event of Moon Disaster*. 2019, MIT Center for Advanced Virtuality.

Fig. 2. Walker Evans. *Child in Backyard*. 1932, The Museum of Modern Art.

Fig. 3. Roger Fenton. *Valley of the Shadow of Death*. 1855, The J. Paul Getty Museum.

Fig. 4-5. Bill Posters and Daniel Howe. Stills from *Veridical Fakes*. 2020, billposters.ch.

Fig. 6. Bill Posters and Daniel Howe. Still from *I wish I knew...* 2019, Instagram.

Fig. 7. Steve Ume and Miles Fisher. Still from @deptomcruise video. 2021, TikTok.

Introduction

This thesis is a close examination of ‘deepfakes,’ which is a form of depiction where animated faces are generated using artificial intelligence (AI) and then incorporated into pre-existing photographically captured video. In technical terms, a ‘deepfake’ is:

[a variety of] methods used to synthesize the appearance of one person to another are largely based on the ability of deep neural networks to learn a representation of multiple facial poses of one face, and transfer that pose to a second face. Underpinning this is a reliance on a large volume of exemplar material which is required in order to successfully train the neural networks to accurately carry out the task. (Lees et. al. 960)

To make a ‘deepfake,’ a person uses an AI algorithm to input a large number of photos and videos of the desired face, ideally taken from multiple angles and lighting conditions. Once the algorithm has been given enough time to identify patterns in the face’s contours, a person can then input a photographically captured video on which to synthesize a new face resembling that of the desired person or character. This technology allows users to put words into someone’s mouth, or to superimpose someone’s face onto a new body and therefore put them into new situations. ‘Deepfakes’ are largely seen as a threat to our understanding of truth, eroding trust in other people and in institutions. Such ability to manipulate the so-called photographic record has led to concerns over its potential harm. Don Fallis, for instance, has raised alarms that ‘deepfake’ use can lead to what he calls an “infocalypse,” whereby we cannot discern real images from fake ones (“The Epistemic Threat of Deepfakes”). Similarly, studies show that while not all viewers will be deceived by a ‘deepfake,’ the very existence of such videos is thought to erode overall trust in news and social media content, meaning that “public debate would become more

difficult, as citizens struggle to reconcile the human tendency to believe visual content with the need to maintain vigilance against manipulative deepfakes” (Vaccari and Chadwick 9).

While I do not want to dismiss the very real concern of widespread lies and mistrust in public institutions, this thesis takes a different approach than Fallis and others. Despite good intentions, such critiques have approached the problem too narrowly, simply equating ‘deepfakes’ with falseness. The complex set of technologies and techniques that are used to create such representations are subsumed under the single tendentious name ‘deepfakes,’ necessarily ascribing an epistemological status in advance. This has led to an oversimplification of the technology and the variety of uses to which it may be put. By contrast, I wish to take a broader philosophic look at what truth is and what its relationship is to *any* technology or medium. The goal of this thesis is to counter the prevailing cultural attitude that any medium or technology is either inherently truthful or inherently deceitful, necessarily true or necessarily ‘fake,’ which such negative critical accounts of this new algorithmic technology assume. While public discourse about this unfortunately named technology is recent and ongoing, these concerns raised by Fallis and others are not new. In fact, these discussions are a continuation of longstanding philosophical questions that new technologies tend to reignite with regard to truth and reality. By taking a more holistic overview of the questions of medium, truth and knowledge, I am not restricting my object of inquiry to ‘deepfakes.’ I pay particularly close attention to photography, for it is the medium against which ‘deepfakes’ are often wrongfully opposed. This is especially pertinent as I argue that ‘deepfakes’ are, in fact, still a photographic medium, for photography is a diverse family of technologies with myriad possible uses.

The very moniker of the technology in question, the ‘deepfake,’ implies inherent falseness. A more neutral term could be “algorithm-generated facial videos superimposed onto

photographically captured videos.” Of course, this is far too wordy to replace what has become such a succinct and popular term. As a result, I rather begrudgingly continue the proliferation of this term, though I place scare quotation marks around the word ‘deepfake,’ unless I am directly citing a source. This is a rhetorical choice to emphasize my argument, that ‘deepfakes’ are not inherently fake, since truth and falseness are actions attributable by humans, and not inherent material qualities. My hope is that this thesis will make clear why such a term does not accurately reflect the metaphysical and epistemological nature of any medium or technology.

To illustrate that all media are equally capable of being used to tell the truth or to lie—and in varying amounts of each—I am now going to provide an example of what we could legitimately call a ‘deepfake,’ created, though, with more familiar means. I hope that I will, in a preliminary way, show you how the popular characterization of ‘deepfakes’ as a particular technology is faulty, confusing matters of discursive choice with notions of technological or material necessity.

“The Search for Truth and Understanding”

The late 1960s was a historical crossroad in the United States, as one turbulent decade neared its end and a new one was on the horizon. In January 1969, Richard Nixon had been inaugurated as the thirty-seventh president of the United States. He had inherited a legacy of deep unrest: his predecessor, Lyndon B. Johnson, had initially assumed office when his own predecessor, John F. Kennedy, had been assassinated in broad daylight. Antiwar protests were ongoing during this period, largely among students and the working class, though “the mass, multifaceted movement that actively fought against [the Vietnam War] eventually comprised Americans across the class spectrum” (Lewis 187). Americans did not just take to the streets to protest the costly, divisive and ultimately unsuccessful war in Southeast Asia; civil rights, Black

power and other self-deterministic and nationalist movements intersected with antiwar activism during this decade (Lewis 15).

Sociopolitical strife was also on a global scale. This was the midst of the Cold War, decades of paranoia and mistrust between the communist Eastern and anti-communist Western Blocs. Historian Matthew W. Dunne underlines the dual fears these tensions brought: “as the specter of nuclear war threatened the safety of Americans’ physical bodies, brainwashing threatened the sanctity of their minds” (4). The rise in patriotism in this era, fuelled by politicians’ stark divisions between the “free world” and foreign communist nations, did not last; the aforementioned home-grown movements of the 1960s were critical of fellow countrymen (Dunne 6-7). No person, and no government, was immune from the mass cynicism of the era.

This period was also the apex of the so-called Space Race, a technological competition between the United States and the Soviet Union, a mirroring of Cold War adversaries. Public discourse has distorted and oversimplified this narrative of space exploration, as Mai’a K. Davis Cross describes. The launch of the first human-made satellite, the Sputnik 1, was made possible largely through the International Geophysical Year, spanning 1957 to 1958 (Cross 1413). Established by the United States and led by the International Council of Scientific Unions, this program was powered by non-state actors, including scientists, teachers and volunteers, from over sixty countries (Cross 1412). It was through this international, largely non-partisan group that humans launched their first satellite in 1957. Rather than creating fear, this breakthrough initially inspired scientists (Cross 1414).

The utopian view of the Sputnik 1 satellite as a vehicle for the betterment for all humans, regardless of nationality, did not last. The Russian scientists who were at the forefront of this endeavour did not want the fanfare of an announcement of the launch, choosing instead to

disclose the news only upon the satellite's success (Cross 1413-1414). This, Cross argues, contributed to the prevailing myth that Sputnik 1 was a secret Soviet project that stoked fear in Americans and propelled governmental resources toward putting the first human on the moon (1414). Despite Sputnik 1 initially being positive news to American scientists, President Johnson and other Democrats had “essentially constructed an alarmist narrative surrounding Sputnik where none had existed before, referring to the United States falling behind... and expressing national humiliation at having ‘lost’” (Cross 1415). Politicians drew parallels to Sputnik 1 and intercontinental ballistic missiles, despite differences between the technologies, using the Cold War disingenuously to push for space exploration (Cross 1415). The National Aeronautics and Space Administration (NASA) had been created in 1958 following the launch of Sputnik 1 for “peaceful purposes for the benefit of all mankind” (“National Aeronautics and Space Act of 1958”). Even with these contradictions—NASA’s apolitical founding mission statement and the nationalistic competitions that spurred the space missions—July 29, 1969, was a momentous day for all of humanity. It was the day that the United States launched three men to the moon.

That evening, Nixon sat down at the Oval Office for a live special report. “Good evening, my fellow Americans,” he began, “fate has ordained that the men who went to the moon to explore in peace will stay on the moon to rest in peace. These brave men, Neil Armstrong and Edwin Aldrin, know that there’s no hope for their recovery.” The camera zoomed in closer on his face as he continued. “These two men are laying down their lives in mankind's most noble goal: the search for truth and understanding.” He alternated his steely gaze between the script in his hands and into the camera, with the knowledge that tens of millions of Americans are listening in awe and horror. In fact, the words coming from his mouth reverberated across the world, sending shockwaves to all and forever altering the course of space exploration.

The United States had tried sending people to the moon, but the return trip resulted in the death of everyone aboard the Apollo Lunar Module *Eagle*. This disaster served as a reminder of the perils of space travel—the ultimate cost of human lives in hopes of learning more about what lies beyond our atmosphere. The tragedy of that day would be forever engrained in public consciousness, a sombre reminder of the dangers of the unforgiving cruelty of outer space. Over the next fifty years, astrophysicists would continue to propose ways of safely returning from the moon, but having witnessed such profound loss, no government on Earth was willing to take that chance again.

Unmasking the Truth

Of course, I lied to you: Richard Nixon did not actually say these words on television, and Armstrong and Aldrin (and Michael Collins) safely returned to Earth after their expedition over fifty years ago. Reader, you probably should have noticed when my writing veered away from the truth. You must have known there was a successful moon landing in 1969, and that this was far from the last time people were sent into outer space. This is what writing, as a medium, allows me to do: I can give you facts, falsehoods and everything else in between, and in whatever configuration I please. Language is an extremely pliable medium, with which I can make many forms of expression, but this is true of all types of depictions. As I said at the outset, we could call this an example of a ‘deepfake,’ if we are to use the term to mean the way a false representation has been created, by manipulating a medium (any medium) at a fundamental level, in this case the medium of language, rather than only the specific medium of video, or moving imagery. I have used facts—the sociopolitical culture of late-1960s America, the Vietnam War, the Cold War, the space race, that Nixon was president—in conjunction with a falsehood, which is that the astronauts failed to return home safely, that they died in space.

One *could* argue that what I did here, and what all ‘deepfakes’ do, is to paste one lie onto a multitude of truths, in the same way that an algorithmically generated face is grafted upon photographically captured video, thus creating an ultimately false depiction. Yet there seems to be more concern about our (new) ability to alter video than the longstanding and inherent ability to alter language, to see more of a threat in pictorial manipulation compared to linguistic manipulation. I do not agree with this assessment, however, for it presumes an inherent truthfulness in video, which I intend to disprove in this thesis. Additionally, this characterization paints all uses of ‘deepfake’ technology—or of any similar technique, such as what I did above in writing—as unilaterally false. Doing so unfairly cancels out the many truths that can be derived from the depiction, while creating an oversimplified impression of what truth is, a matter that I will explore in depth in Chapter 2. For now, it will suffice to say that this technique of ‘deepfakery’ is possible with any medium and is not an inherent quality of so-called ‘deepfake’ technology: all representations created with any medium need to be assessed for truthfulness.

There are numerous ways you can verify the claim that the Apollo 11 mission successfully put humans on the moon, and that they returned safely, from testimony of those intimately involved in the space mission, as well as from the countless media reports (newspaper articles, television broadcasts, photographs etc.) of the period. There is an enormous amount of information supporting the fact that there was a successful trip to the moon in 1969, so much that, despite not having been alive at the time, I have confidence that these basic facts are correct.

In my linguistic ‘deepfake,’ I have provided citations where necessary, allowing readers to determine the veracity of (some of) my claims, to establish (a sense of) the truthfulness of my written ‘deepfake.’ These sources, like all forms of second-hand information, must be judged on their own merits: the credentials of the author, the reputation of the publisher, the date of

publication, the author's stated (or implied) intent, the sources cited, and so on. What I could not provide, though, were citations of references to corroborate my false claim that the astronauts died, and that Nixon announced their deaths on television. Our understanding of truth necessarily relies on information we take in, within such complex discursive and institutional contexts, either through direct experience or from other first- and second-hand accounts of other people.

This action of determining truthfulness—which I will refer to more specifically as judgments—forms the basis of my argument with regards to 'deepfake' technology. Consciously and subconsciously, we are constantly assessing what information to take as trustworthy, whether through our direct perceptions or from what others convey to us from their own perceptions. Further complicating our search for truth is the simple fact that people can lie or unintentionally tell you a falsehood. Additionally, not all information is given with the express purpose of being factual, and judging what is true or untrue is complicated at best. More often than not, the true and the false are deeply intermingled, as is the case with my account of Apollo 11.

Same Spiel, Different Medium

The specific details of my linguistic 'deepfake' are not chosen randomly. There is indeed a video circulating of Nixon addressing the nation to announce the failure of the Apollo 11 mission to the moon and the deaths of the astronauts. It is, however, a 'deepfake' video developed by the MIT Center for Advanced Virtuality. I argue that what the video's creators are doing is effectively no different than what I have done above in writing—the only main difference is the medium and how it was created. Like my written account of a moon disaster, this video melds the true and the false. Nixon's speechwriter William Safire had drafted an alternative speech shortly before the launch, aptly titled "In Event of Moon Disaster." The words

Nixon appears to be saying in the video therefore are partially based on history, even if that is not a video of Nixon saying these exact words in front of a camera in 1969.



Fig. 1. Francesca Panetta and Halsey Burgund. Still from *In Event of Moon Disaster*. 2019, MIT Center for Advanced Virtuality.

To make this video, the researchers fed a large selection of videos of people's faces talking into an artificial intelligence (AI) model, following a specific algorithm. Using what is known as machine 'learning,' the model looked for patterns in talking faces (Panetta and Burgund, "Behind the Scenes"). Once the algorithm has been fed enough videos to sufficiently differentiate parts of the human face talking, it can analyze inputted videos to synthesize moving faces through pattern recognition. Next, the researchers recorded an actor reciting the speech. They then needed a video on which the algorithm could synthesize a face; they selected Nixon's

televised resignation speech because its somber tone matched their desired effect (Panetta and Burgund, “Behind the Scenes”).

These two videos—the actor reciting the speech and Nixon’s resignation—are inputted into the same algorithm, which is able to map the actor’s face onto the target video and synthesize Nixon’s mouth to match the movement of words. While other ‘deepfakes’ will visually synthesize the full head, this example only alters the mouth, leaving the head movements, blinks and twitches identical to that in the resignation speech (Panetta and Burgund, “Behind the Scenes”).

A similar process was used to create Nixon’s distinct voice. Researchers recorded hundreds of short clips of the actor reading Nixon speeches about the Vietnam War (Panetta and Burgund, “Behind the Scenes”). These recordings—the original Nixon audio and the actor’s mimicry—once inputted into the model, “enabled the AI to learn how to ‘translate’ one voice into the other” (Panetta and Burgund, “Behind the Scenes”). The algorithm established the patterns between Nixon’s and the actor’s distinct vocal qualities, allowing it to construct audio of Nixon to match the actor’s “moon disaster” speech recitation.

To make it feel even more authentic, the researchers preceded the ‘deepfake’ clip with a few minutes of archival video of the Apollo 11 launch, just as I provided initial factual information in my account. They included shots of families huddled around the television, the rocket blast-off, and even some clips of the astronauts, selectively edited to appear as though they were losing control of the spacecraft, cutting to the live special report from President Nixon for greater dramatic effect. As I will discuss in Chapter 2, these aspects of the video are examples of what Britt Paris and Joan Donovan call “cheap fakes,” where it is simpler, and often more effective, to manipulate already existing media for deception, often simply through editing

video or film footage, rather than going through the efforts of making a more sophisticated depiction, a wholly new ‘deepfake’ (2-3).

The Rise of ‘Deepfakes’

Scholars cite 2017 as the year in which ‘deepfakes’ first began to spread online. Though similar technology existed and was in use professionally,¹ this was the year in which amateur use, and examples online, proliferated. The term was most likely coined by a Reddit user, identified as “u/deepfakes,” who began posting such videos that year. It is presumed that the term is a portmanteau of “deep learning,” an algorithm technique, and “fakes.” In autumn 2017, the #MeToo movement against sexual assault was started, coinciding with when u/deepfakes began posting explicit videos of porn stars with AI-generated faces of actresses, such as Daisy Ridley and Gal Gadot. Shortly after, once u/deepfakes’ creations gained notoriety online, journalist Samantha Cole posted the provocatively titled article “AI-Assisted Fake Porn Is Here and We’re All Fucked” on *Motherboard* (*Vice*’s technology section). Here, Cole interviews u/deepfakes, though even she admits she did not determine his real identity. (u/deepfakes’ gender is unknown, though Cole uses the pronouns “he” and “him” to refer to this user.) For u/deepfakes, it is futile to suppress emerging technology, even if it can have nefarious uses (Cole). As he tells Cole, “every technology can be used with bad motivations, and it’s impossible to stop that.”

u/deepfakes’ online fame was short lived: a mere four months after Cole’s article was published, Gfycat (a GIF-hosting site), Pornhub, Twitter and Reddit each announced bans on pornographic ‘deepfakes,’ all within the span of a few days (Hern; Kelion; Robertson). While

¹ As outlined in Chapter 3.

sexually explicit versions did not last long on Reddit before being banned, and the user is no longer on Reddit, the technology, whose moniker he coined, has flourished across the internet.

It is through online circulation that I first became aware of ‘deepfakes.’ The President Nixon moon-disaster ‘deepfake,’ and many others before it, spurred my interest in the technology, its capabilities, and its infamy. Throughout the thesis, I will refer to other contemporary examples, each with their own unique set of circumstances that show the complexity of ‘deepfakes,’ a complexity that cannot be oversimplified by simply describing it as an inevitably dangerous technology. Like the Nixon video, a plethora of ‘deepfakes’ have been made specifically to educate or warn us of the dangers of this technology, whether it is journalist Claire Wardle posing as the singer Adele for *The New York Times* or Jordan Peele as President Obama (*BuzzFeedVideo*). From my own research, these didactic examples—in addition to more comedic ones, whose humour derives from our awareness that it is a ‘deepfake’—far outweigh the number of ‘deepfakes’ that have been identified as potentially causing harm to our conceptions of truth. To create a ‘deepfake,’ in fact, takes considerable time and resources, provided by MIT and other institutions for the creation of the now familiar didactic examples. The expense and complexity of this technology has limited the creation of more deliberately false examples. That is not to say we should not beware of potential misuse, as cost and complexity potentially decrease; rather, we need to understand that this technology, like any other, can be weaponized for harm just as much as it can be a neutral or even positive tool.

My Argument

Like the ‘deepfake’ President Nixon says, I, too, am on “the search for truth and understanding.” As with any new technology, ‘deepfakes’ have come under intense scrutiny for their potential weaponization for misinformation. In addition to Cole’s anxiety-inducing

Motherboard article from 2017, there have been countless other alarmist headlines, such as: “Deepfakes Are Going To Wreak Havoc On Society. We Are Not Prepared” (Toews); “The Scary Truth Behind The FBI Warning: Deepfake Fraud Is Here And It’s Serious—We Are Not Prepared” (Gow); and “Remember the ‘deepfake cheerleader mom’? Prosecutors now admit they can’t prove fake-video claims” (Harwell). A study of English-language reports about ‘deepfakes’ published between January 2018 and July 2019 were found to be “highly speculative, and largely painted a dystopian picture of the consequences of the technology,” with stories seldom reflecting its immense creative storytelling capabilities (Wahl-Jorgensen and Carlson 809). The study authors note that this can be partially attributed to the timeliness of news stories, which were written in the wake of negative examples of ‘deepfakes,’ and to the fact that “the specter of deepfakes provides a rallying point for journalists to reassert their social value in both implicit and explicit ways” (817).

While these are specifically examples of journalistic concern, this paranoid discourse is far-reaching across audiences, given, in part, the news media’s role in the dissemination of ideas. For instance, political communication professors Cristian Vaccari and Andrew Chadwick state that traditional tactics for combatting online misinformation, by “encouraging the public to seek out alternative sources of information and to juxtapose these with any utterance or source that claims to be authoritative,” may no longer be tenable with the rise of ‘deepfakes’ (9). They attribute this to our propensity to trust that “whatever is said in public, a real person has said it, even though the statement may be false,” as well as the difficulty in proving that a video is not a ‘deepfake’ (9).

In the face of such alarming accounts, I want to emphasize more careful philosophical descriptions of the technology. I also want to look at technology as a whole, looking for common

threads in the many concerns expressed about ‘deepfakes.’ I do not want to downplay the very serious threats to knowledge, and thus society, of any deception. Instead, I want to question why we are targeting the ‘deepfake’ technology itself as the source of mis- and disinformation, instead of those individuals who spread lies by any means at their disposal.

Despite these very real concerns, the blame is too often placed on ‘deepfakes,’ not on those with malicious intent. Put simply, deception using videos, as with any other medium, is a moral problem, not a technological problem. Truth or falsity are not inherent to *any* representation. Truth is a value that humans ascribe to expressions, or what philosopher Michael Lynch calls a “thick value,” emphasizing the complexity of the concept, but also the role individuals play in judging truth and falsity.² Likewise, truth is not fixed. I can show you a photo I took this morning of my sunny window and say, “look at how sunny it was this morning,” but I can also show you the same photo a few days later and say the same thing, despite it having rained that morning. In the first example, I am telling the truth; in the second, I am not. In both cases, though, the image itself has not changed. A photograph of a sunny window, on its own, contains neither any truths nor falsities. This is the same for any representation, whether it be written, drawn, painted, sculpted or photographed. Context is also the key to the truthfulness of my statements, which is independent of the image itself. In the above example, the time in which I tell you to “look at how sunny it was this morning” is critical to whether or not I am telling the

² Lynch says that “truth, like courage or keeping a promise, is something philosophers call a thick sort of value; it has normative *and* nonnormative aspects . . . [W]hen we say that a belief is true, we are at once evaluating it—saying it is correct, good, worthy and so on, *and* describing it as portraying the world as it is” (19-20). Lynch rejects an extreme relativism on matters of truth, but calls for a relativism without nihilism, “some aspect of relativism without slipping into nihilism” (41). See Lynch, Michael P. *True to Life: Why Truth Matters*. MIT Press, 2004. For a similar account of a moderate relativism about truth, see Blackburn, Simon. *Truth: A Guide*. Oxford University Press, 2005.

truth. Nonetheless, there is the persistent impression that photography is inherently more truthful, or more representative of reality, and that ‘deepfakes’ are inherently more deceptive.

Chapter 1 will address the misconception, both implicit and explicit, that any medium or technology is more truthful, or more prone to falsities, than any other. This is perhaps best exemplified by how Western society has championed photographs as inherently truthful. On the contrary, digital media, including ‘deepfakes,’ are often characterized as inherently false. We see this dichotomy emerge when discussing potential mandates to label AI content, which will serve as an entry point to this chapter. I follow Patrick Maynard’s assertion that our lack of attention to the philosophical implications of photography as a technology has led to this misplaced trust in photography and emerging types thereof (97). Maynard uses photography to study technology as a whole, thus avoiding an overemphasis on photography’s supposedly unique characteristics, which leads to “not only an inability to help us understand new examples as they appear, but also a tendency to isolate photography from related, even closely related, phenomena” (17). I argue that ‘deepfakes’ should be understood as a primarily photographic medium, despite the presence of computer generation, as a phenomenon “closely related” to photography.

Having established that truth and falsity are not located in the medium or technology, I continue my inquiry in Chapter 2 by considering where these value judgments are located. I use traditional philosophical approaches to truth and the nature of reality to differentiate two types of truth. While one of these types of truth is more useful for determining truthfulness, we mistakenly conflate this truth with the other, more abstract type, leading to undue confusion and defeatism. I also consider what it means to lie, which, much like truth claims, are hard to determine because we do not have access to other people’s thoughts and perceptions. Given these ambiguities, I look to contextual clues for better situating our critical thinking.

Finally, in Chapter 3, I expand upon the fears of AI and technology generally, which has been briefly discussed in Chapter 1. I describe the development of technologies that were foundational in the development of ‘deepfakes,’ which demonstrates our long history of using computers to alter photographed faces. I then discuss how AI, and algorithms in particular, have been wrongfully characterized as wholly autonomous processes, independent of human agency. Finally, I look at the different ways in which we define digital and analog, opting for a philosophical approach, as well as making an argument against overly simplistic material determinism.

The emphasis of my thesis is on truth claims and ‘deepfakes,’ not on artistic merits of this technology. That said, I cannot completely dispense with all notions of art, even with regards to these media objects, for there is a deep connection between the distinctions of truth and falsity, fiction and nonfiction, and the notion of art. This is exemplified in my above account of the moon landing failure. Thus, while this is not an inquiry strictly rooted in the realm of the arts, aesthetics informs a great deal of my research, which, in turn, can be applied to the art world. While undeniably linear, by virtue of being the written word, this thesis is somewhat circular. The philosophical concepts in which I am engaged are so complexly intertwined that a strictly linear account is unattainable. Throughout the thesis, I refer backwards, as well as offer a glimmer of concepts I expand upon in subsequent chapters. In the conclusion, I shore up these discussions of truth and technology, as well as provide an account of this technology where it can be used positively and even for telling truths. I also lay bare the unknown variables with my research, for this is a very recent and quickly evolving technology and area of study. Finally, I briefly show where this argument fits in with larger concerns humanity has been grappling with over time.

Chapter 1: Separating Truth from Technology and the Medium

‘Deepfakes’ have been the cause of acute fears surrounding misinformation online, leading to some extreme proposed measures to curb their use and spread. One measure that researchers have suggested is the use of labels warning viewers of AI use. In the first of a two-part series for the research network *First Draft*, the authors ask, “how should we label AI media in ways that people understand? And how might the labels backfire?” (Shane et al, “From deepfakes to TikTok filters”). The researchers who penned the piece have combined knowledge in misinformation, user experience and internet policy. The article provides thirteen possible methods for disclosing content made (in part or in full) with AI, with explicit references to ‘deepfakes.’ The methods are organized into four broad categories: “Labels: Covering the media”; “Metadata: Describing the media”; “Interruption: During the media”; and “Annotation: highlighting the media” (Shane et al, “From deepfakes to TikTok filters”). The authors identify the strengths and weaknesses associated with each method of disclosure. For instance, a watermark on an AI image will be retained in screenshots, though removal with Photoshop or a similar software is always possible. Other examples of labels are to abruptly cut to a statement midway through the video, and to use side-by-side ‘before and after’ images to illustrate the difference made by AI.

It is in the second and final article that the authors ask moral questions that are lacking in the first article. They contend with the issues of when and how to apply a warning label, and how to avoid “[inspiring] false confidence that something is fabricated when it isn’t, or [suggesting] that all unlabeled content is trustworthy” (Shane et al, “There are lots of ways to label AI content”). While worthwhile ethical questions generally, it is troubling that they only answer how, when and where to use these labels, but not the most important question of all: *why impose*

these labels in the first place? They ask why certain content should receive certain types of labels, but they assume that labels are necessary, or even inevitable.

These articles assume that to be ethical, all use of AI must be disclosed. The goals of the publication, *First Draft*, provide additional context to support this underlying assumption: “our mission is to empower society with the knowledge, understanding, and tools needed to outsmart false and misleading information” (“About”). This mission statement accurately reflects the importance of empowering society with knowledge and understanding, but its inclusion of tools is potentially troubling. Using the term in the broadest sense, the assessment of truthfulness has always been an individual endeavour, since truth, or the lack thereof, is not inherent to any object, but rather a judgment made by people using their mental faculties. The issue of people being misled by ‘deepfakes’ and other technologies is a moral, not merely a technological, problem. This is not to imply that technology is a neutral fact in the world. While every technology has its specific qualities and inherent limitations by virtue of its materials, they all require moral appeals in how it is used and shared. It is also the case that new technologies often disrupt existing social systems and institutional structures for upholding truthfulness. However, I argue against technological determinism, as humans play a central role in the creation and use of technologies, such that we have free will to govern our use. We could consider knowledge and understanding to be tools, but that *First Draft* lists it as a separate method suggests a more literal definition, including these labels that the authors describe. As I argue throughout my thesis, we have always had the same means available for judging if someone is lying.³ Such types of

³ There is great historical, cultural and physiological diversity, among other factors, that make each person’s circumstances unique. These differences influence how we judge truthfulness, but, as I elaborate on in Chapter 2, such instances of evaluating truth claims are highly specific to the individual or body of people to whom the evaluation is being made, not any representation of Truth more broadly.

regulation or stifling of a particular technology are ineffective methods for curtailing malicious use or misinformation, with potential dangers.

The implicit rationale for urging disclosure is that there is something fundamentally different about representations created using AI, that AI is somehow ontologically distinct from photography and other technologies or media. The need for a label is also based on the presumption that viewers are not able to discern for themselves if AI was used, if such a label were not applied. I am not suggesting that we cannot be harmed by what we do not know, but there is something odd in the presumption that representations created with AI are inherently harmful. Needing a label also implies that we would not be able to detect the use of AI otherwise, and that failing to do so would have negative consequences. This is effectively equating AI with falsities. In the most generous interpretation, this singling out of AI demonstrates that we make a correlation between method of creation and truthfulness. That we perceive AI to be deceptive leads to our own distrust of the technology and the images created with it.

While these researchers have a relatively nuanced look at when, where and how to use these labels—methods that might be useful in very specific contexts as an ethical choice—I want to take a different approach to this question. Rather than beginning with the logistics of labelling, followed by the ethical implications, I want to start with a broader philosophical discussion of this practice. Thus, in this chapter I will contend with the persistent fears that surround *any* new technology. Following this history of cultural attitudes will demonstrate that current fears of ‘deepfakes’ are nothing new and are, in fact, all but inevitable. In addition to its newness, the distrust of ‘deepfakes’ is also due to a general collective wariness towards AI. However, this chapter is a starting point for understanding technology; it is in the third chapter where most of

this AI-specific discourse will resurface. Depending on the person, the autonomy of AI can be seen as either a worry or a cause for celebration. This autonomy is a large part of why people are so fearful of AI in general, though this is a recurring theme in discourse on technology more broadly.

The fundamental distrust of ‘deepfakes’ is part of a broader skepticism. This skepticism belongs to a larger school of philosophical thought, whose origins in the Western written tradition can be traced back to Plato (Landesman and Meeks 9). Skeptics say that our subjective, perceptual experience of the world prevents us from ever truly knowing reality. Since we do not know for certain if anything we see, hear, touch, taste or smell is reality or a deception, according to this philosophy, then we do not have access to this greater realm of reality. In its most basic form, skepticism is “a general attitude where no belief is immune from a call for justification,” requiring a more rigorous self-reflection on the nature of knowledge and doubt (Landesman and Meeks 9). The notion of constantly interrogating what you know can be taken to extreme measures, leading you to conclude that this fallibility renders the pursuit of knowledge fruitless. This is not what Plato seems to be saying; rather, via dialogue, “Socrates concludes that his wisdom lies in his knowledge of the limits of human wisdom, that he—unlike his fellow Athenians—did not claim to know what he in fact did not know” (9). Nonetheless, skepticism, in the popular vernacular, has transformed into a broader concept denoting mistrust, including of the images you see. Regardless of any one skeptic’s attitude toward the pursuit of knowledge, skepticism, both in its original Platonic form and in its modern colloquial use, gestures at our uneasy relationship between reality and knowledge. I will return to this school of thought in Chapter 2, where I argue that skepticism overly complicates our understanding of truth, insofar as we use it for evaluating truth claims.

Throughout this thesis, I will contend with media theorists whose work, whether they state it or not, follows in the tradition of skepticism. One such theorist is D.N. Rodowick, who says that the move towards digital film “may [have indicated] a philosophical retreat from the problem of skepticism to an acceptance of skepticism” (175). In other words, the introduction of digitally transmitted electronic media has ‘proven’ that we are not to trust what we see, what we perceive. Implicitly, this means that Rodowick sees digital media as more manipulable, or as inherently less trustful than other media, such as photography. This is the same approach that the *First Draft* authors take. Rodowick is following in the footsteps of Stanley Cavell, who famously argued that “the reality in a photograph is present to me while I am not present to it; and a world I know, and see, but to which I am nevertheless not present” (23). Photography, according to Cavell, makes manifest our fundamentally phenomenological experience of the world, thus corroborating that we are limited to our own perceptions.⁴

Other scholars have offered more foreboding accounts of skepticism and photography. Hany Farid, a UC Berkeley professor who specializes in digital images, suggests that photographically-generated videos offer objective access to past events that have been captured on camera. As part of a multimedia feature on ‘deepfakes,’ he warns CNN: “If we can't believe the videos, the audios, the image, the information that is gleaned from around the world, that is a serious national security risk. We have to get a handle on how to authenticate this media” (O’Sullivan). I agree with Farid that a collective agreement on certain truths is important to national security, but I want to stress that it is not the images themselves on which we should focus our trust, but rather the values humans attribute to them. I also disagree with media

⁴ A great deal has been said about Cavell’s theories on film. For instance, see Rothman, William, and Marian Keane. *Reading Cavell’s The World Viewed: A Philosophical Perspective on Film*. Wayne State University Press, 2000.

authentication being crucial to combatting misinformation. As will be explained in this chapter, all media are susceptible to misuse, but I will explain in Chapter 2 that we use the same means of judging truth claims, regardless of the complexity of the process for making any particular representation.

Farid’s characterization, whether intentional or not, contrasts an inherent falseness of AI-generated representations with the supposed objectivity of photographic images. Because it is possible to make a ‘deepfake’ video of a politician saying something inflammatory, so goes this reasoning, all videos must be treated as potential ‘deepfakes,’ and thus we must now be extra skeptical. This skepticism is thought to prevent blind faith in such ‘fake’ videos, but it could also lead to a distrust of videos altogether, both with harmful outcomes to democracy, as the authors of the *First Draft* articles even admit (Shane et al, “There are lots of ways to label AI content”).

This human detachment from reality is an underlying theme of those who say we are living in a post-photographic epoch. It is this movement that Rodowick, writing in 2007, appears to be echoing. William J. Mitchell, a prominent scholar in the post-photography movement, had written in 1992 that “digital images are, in fact, much more susceptible to alteration than photographs...or *any* other kinds of images” (7, emphasis in original).⁵ Despite the popularity of this movement, I argue against the notion of living in a post-photographic world, in favour of continued allegiance to our human perceptions, however subjective they may be. This means that both analog and digital photography can be easily used for lying, as I will explore further in Chapter 2.

As I intend to make clear in this thesis, all media are equally capable of manipulation. I also want to challenge our understandings of truth, since photography has often been wrongly

⁵ I argue against digital media containing less information, and thus being more manipulable, in Chapter 3.

treated as inherently truthful. In the next chapter, I will explain how we have always had the same information available to us for making truth judgments as we do now. Truth and falsity, insofar as we use these concepts in our daily life, are statements we can ascribe to objects, not part of the objects themselves. I can use the same photograph to tell the truth or to tell a lie. I can even say something that I think is true, but which is not. In all of these instances, the object does not change. Likewise, I can tell a truth or a lie using any form of depiction I please. I am confident in saying that there has never been a medium that has not been used for lying. As I demonstrated at the outset of this thesis, I can use words to tell the truth, but I can also use words for deception.

Despite our long history of deception using imagery, there is a prevailing assumption in Western culture that some *types* of depiction are more truthful, or more accurate at representing reality, than others. This is especially true of photography, which has been simultaneously praised and disparaged for its level of automatism since its inception in the nineteenth century. Photographic imagery, and pre-digital video, were themselves never the arbiters of truth; ‘deepfakes,’ and other such digitally manipulable media, are no different. Indeed, technology and medium are interconnected concepts that shape our understanding of depictions.

I will begin my enquiry of ‘deepfakes’ as a tool for deception by first considering technology as a whole, and by considering representation—and specifically photographic representation—as an instance of technology. To do this, I will look to the defining characteristics of technology that Patrick Maynard outlines in *The Engine of Visualization: Thinking through Photography*. Though published in 1997, Maynard’s account of technological change still resonates today, whereas emphasizing the photographic image over the technology

“falls short in its ability to help us understand *new* kinds of photo imagery as they appear” (17, emphasis in original).

‘Deepfakes’ are not the first imaging technology to call into question our understanding of technology; as historian Robert D. Friedel puts it, “nothing did more than photography, for example, to blur the line between science and art, and to identify technology and invention as mediating instruments between these realms” (341). I want to draw our attention to the history of technological change, in order to see the ways in which ‘deepfakes’ are a continuation of the same discourses. This is in line with the crux of Maynard’s book, that “what we call ‘photography’ is a *technology*” or, rather, a family of technologies (3, emphasis in original). While film scholars frequently acknowledge the technological history of photography—the various purported forefathers, the techniques used, and the innovations that followed— “such histories are nearly devoid of the reflection upon the nature and history of technologies generally and never seem to bring insights there to bear on their subject matter” (Maynard 16). I follow Maynard in prioritizing technology over the image itself, with regards to epistemological and metaphysical inquiry. Doing so provides a more nuanced look at what images are as a whole, demonstrating the characteristics which photography and ‘deepfakes’ share with all other modes of representation.

Employing Maynard’s technology-first account of photography also shows us the ways in which photography, a family of technologies that now includes ‘deepfakes,’ have two functions shared across all photographic images: depiction and detection. Understanding how we use photographs, based on these two functions, prevents us from making two major assumptions in photography. Firstly, it prevents us from assuming that photography is an entirely depictive technology. Secondly, it demystifies the detective quality of photography, which is what has led

to an overconfidence in its supposedly inherent truthfulness. I use these two functions, as well as a brief comparison between photography and AI, to show our paradoxical relationship between automation and trustworthiness, or lack thereof.

Whether or not photographic technology inevitably leads to a truthful medium—or AI to a deceptive medium—is at the heart of this debate. The concept of medium is complex, leading to much debate into its function in our understanding of depictions. To further illustrate the interwoven relationship between medium and technology, I will refer to John Guillory’s essay “Genesis of the Media Concept.” Using his account of the conceptual effects of new “technical instruments,” the new electric technologies of communication and representation that emerged in the nineteenth century, I will build upon his argument that “the development of new technical media perplexed thereafter the relation between the traditional arts and media of any kind” (Guillory 321-322). When Guillory uses the term “new media,” he is specifically referring to the telegraph and phonograph, which were once so novel, and whose novelty is manifested in all subsequent communication technologies. These inventions ushered in the modern age of technology, where we first became concerned with the question of communication (Guillory 321).⁶ However, the term “new media” is usually applied to more contemporary technologies. In a broader sense, new media is defined as “new means of mass communication considered collectively,” which would include today’s latest technologies, especially digital technologies (“New Media, n.”). Though Guillory writes about nineteenth-century ‘new’ media, the points he makes nonetheless apply to new media today.

⁶ John Durham Peters has also described the responses to new technologies in the nineteenth century, and the effects they have had on our concepts of communication and of truth, contributing to the emergence of what he calls “the problem of communication.” See Durham Peters, John. *Speaking into the Air: A History of the Idea of Communication*. University of Chicago Press, 1999. See also Marvin, Carolyn. *When Old Technologies Were New: Thinking About Electrical Communication in the Late Nineteenth Century*. Oxford University Press, 1988.

This essay by Guillory was published in 2010, before the popularization of ‘deepfakes,’ but we can still use his theoretical framework to demonstrate that such technology does not signal an abrupt departure from our current understandings, but instead a continuous re-examination of the meaning of a medium and—through their deeply entwined relationship—of technology in general. By taking a historical look at how we have characterized technologies and communication through the centuries, we begin to see recurring patterns, which signify that photography, and more recent technologies, are not a radical departure from our understandings of what constitutes a medium. Moreover, the relationship between truth and *any* medium or technology is not straightforward.

Technology

The topic of technology is most often invoked in popular discourse under the umbrella of “new media,” when discussing computer-based images created and/or displayed on digital screens. This is understandable, given its obvious presence in the form of pixels and chips, but it is unfortunate that theories of technology are left implicit in relation to even the ‘simplest’ technological medium. A pencil, for one, is a piece of technology, but accounts of media seldom consider the ways in which they are similar to, say, computers, opting instead to take for granted what it even means to call something “technology.” For Maynard, this misunderstanding of technology, or a lack of broader inquiry into the term, is “one of the main reasons why we have been in such a muddle about the arts for so long” (97). We must not think of art and technology as contrasting concepts.⁷ This thesis follows in the steps of scholars like Maynard, who emphasizes the importance of technology in understanding images.

⁷ Likewise, we must not treat art as subordinate to science, nor should we try to apply scientific processes to the artistic process. I elaborate on this in Chapter 3.

But what is technology? I follow Maynard’s simple definition of technologies, that they are the “extenders or *amplifiers of our powers* to do things” (75, emphasis in original). Robert D. Friedel offers further qualifiers to better situate us. His definition emphasizes the cultural dimensions through which we create and use technologies: “the knowledge and instruments that humans use to accomplish the purposes of life” (Friedel 1). Technology, in its material form, cannot be wholly biological or mental; however, it is closely related, as everything is reducibly derived from nature, and our mental faculties are required to imagine, create and use technologies (Friedel 1). That technology exists is because of our desire to enhance our abilities to do something. I could use my finger to mark an image in the sand, or I can use any number of tools created to enhance my ability to make markings. While I agree with Maynard’s definition, that technology amplifies our power to do things, I want to stress the importance of our mental processes in using technology, for it is crucial in understanding *why* and *how* we use technology, which is equally important in understanding *what* technology is.

It is because we want to amplify our powers to do things that we create and use technology. Maynard stresses that these are *physical* amplifiers, from which we can make value judgments about what we want to do with the technology (78). Before ascribing values to the technology, it is important to look at what technologies do on a fundamental level. Doing so can prevent us from equating the technologies themselves with any human-defined purpose, which, for example, could lead us “to generalize...the nature of something called ‘the photograph’” (Maynard 17).⁸ We should think of amplification as manifested in a primarily physical capacity; in the case of photography, this technology has amplified our ability to make markings on

⁸ As I describe later on in this chapter, this generalization is commonly manifested in medium essentialism, whereby theorists prescribe certain uses to ‘optimize’ a medium’s supposedly inherent qualities.

surfaces (Maynard 79). It is through this amplified surface marking power that photography amplifies, for example, our ability to imagine. There are many types and number of amplifiers possible. Possible amplifiers of photo technologies generally are: the speed of creation, the level of detail possible, or the portability of the camera. Specific to any camera, Maynard points to exposure length and aperture size as potential amplifiers (79). These amplifiers are not values, but we can use these potential amplifiers to make value judgments.⁹ For example, using a telephoto lens amplifies your ability to take a picture from far away, which is a good value if you want to photograph a bird in a distant tree. Lenses are not sentient like humans, so they do not value that they amplify our ability to shoot objects from afar.

It is not just amplification that is possible with any technology, for “we must bear in mind the seeming paradox that just as often, our powers to do things are amplified by physical *filters* or *suppressors*” (Maynard 78, emphasis in original). Similar to Newton’s third law of physics, all amplifiers must be accompanied by an equal opposite filtering effect. While a camera can allow for more autonomy in creating a perceptually realistic image, for example, the author has less control over the details of the image. As I will explain below, the limits of a photograph, as manifested by the technology used, do not mean we can only use photography for certain kinds of depictions and not others. However, there are many influential accounts of technology that assume certain values are inherent in particular technologies. This is most keenly seen in theorists who ascribe specific uses for a medium (as Siegfried Kracauer does with photography¹⁰), but we can see this presumption of inherent value in technology more broadly.

⁹ I go into more detail on what values are in Chapter 2.

¹⁰ I go in more depth on Kracauer’s essentialist views on the use of photography later in this chapter.

Martin Heidegger is one such philosopher whose account of technology is more value driven than that of Maynard. While Maynard emphasizes technology as an action, Heidegger looks to the purpose, or end goal, of its use. For him, technology is a human process that allows us to see truths of the world (Heidegger 12). Heidegger refers to the *Gestell*, or enframing, of a technology as the “technological framework [that] is inherently expansionist and can reveal only by reduction” (Krell 309). Heidegger thus sees reality, or the natural world, as the bearer of truth, and technology as our way of revealing the world to ourselves (12). Though this may seem to be a positive outlook, Heidegger is very cautious: “What is dangerous is not technology. There is no demonry of technology, but rather there is the mystery of its essence. The essence of technology, as a destining of its revealing, is the danger” (28). Heidegger is wary of what will happen if we overemphasize this revealing, or if we reveal too much for humanity to handle. He continues:

The question concerning technology is the question concerning the constellation in which revealing and concealing, in which the coming to presence of truth, comes to pass. But what help is it to us to look into the constellation of truth? We look into the danger and see the growth of the saving power. How can this happen? Here are now and in little things, that we may foster the saving power in its increase. This includes holding always before our eyes the extreme danger. (Heidegger 33)

The issue that I have with Heidegger’s account is that he posits a kind of essentialist account of technology, whereas Maynard’s account is more functional. Unlike Heidegger, I do not see technology as having an inherent value, either good or bad; rather, technology is a discourse that necessarily relies on human cognition to derive its value.

Focus on Photography, Not Photographs

Having started by discussing the technology, it should be clear that I am focusing on *photography* over photographs. Despite historical accounts that emphasize its technological history, the philosophical study of photography focuses primarily on the *photograph* (Maynard 16).¹¹ Maynard proposes we do the opposite: studying photography primarily as *a family of technologies*, with photographs defined as “the physical marking of surfaces through the agency of light and similar radiations” (3). When we study photography mainly through its resulting images, there is the tendency to overlook the many purposes for which this family of technologies may be used. Not only do we ignore the range of photographic processes—different chemicals, surfaces, and camera apparatuses—but also the kinds of photographs that are made, their purposes and uses. We cannot predict the form and number of uses possible, but here are a few common uses of photographs: to remember something, to spread news, to advertise products, to study natural phenomena, to make medical diagnoses, to tell stories, and to provoke thought (Maynard 16-17). Given this diversity of uses, there is no practical answer to what an ‘ideal’ photograph is. Similarly, many other technologies can be used for any of the above uses, meaning that they are not uses unique to photography. Rather than seeking an essential definition, it is more fruitful to look at photographs by what they do all have in common: *photography*, the technologies and processes through which a photograph is created.

¹¹ This is not to say that scholars have not thought about how they are made. Rather, the biggest problem with such photograph-first accounts is that they are based in inaccurate assumptions about the act of taking a photograph. Roger Scruton, for instance, overemphasizes the relationship between referent and photograph, stating that “the ideal photograph... stands in a causal relation to its subject and ‘represents’ its subject by reproducing its appearance” (587). As I will argue later in this chapter, such talks of ‘ideal’ photographs are irrelevant to understanding how people use this technology. See Scruton, Roger. “Photography and Representation.” *Critical Inquiry* vol. 7, no. 3, 1981, pp. 577-603.

By contrast, D.N. Rodowick’s theories of photography hinge on the understanding that “fundamentally, a photograph is an automatic transcription of a past state of affairs” (78). He is correct that temporality is an element of a photograph (given that they are created through the detection of light in a particular timeframe), but he is mistaken in his assertion that, regardless of intended use, “before transforming [photographs] as signs, fictions, of works of art, we approach them as historical documents” (78). While we may acknowledge that a photograph in an advertisement was created when light marked a surface at a particular moment in time, in what way does the photograph *represent* this? How does, say, a cheeseburger on a billboard call attention to itself as “a historical document”? If anything, the photograph calls attention to *the future*, namely in that it beckons you to buy and eat a cheeseburger. This example demonstrates that it is not inevitable for us to treat photographs as “an automatic transcription of a past state of affairs” because that is a value that humans discriminately ascribe to particular photographs. Such a statement undermines both the involvement of the person in creating the photograph, but also gives the humanlike quality of agency to a nonhuman entity (i.e., the light transcribing), whereas agency is an action that requires intention and purpose. Rodowick’s work also emphasizes the photograph over photography, though he is also fixated on the process of photography, insofar as he uses it to give photographs ontological properties distinct from other types of picture-making. As I will describe later in this chapter, photography does not have any special ‘essence’ that mandates how we should use it optimally. This misconception arises from overprioritizing one of the two functions of photography.

Primary Functions and Markings

Rather than values or purposes, which are exclusively ascribed by humans, functions are descriptions of what any particular technology (or family thereof) *does*, described neutrally.

Maynard identifies two primary functions that are universal to all photographic technologies: depiction and detection. Though photography is not the only family of technologies to harness these two functions, “‘photography’ might be most simply characterized as *the site of historically the most spectacular interaction of depictive and detective functions*” (Maynard 120, emphasis in original). These two functions—depiction and detection—are deeply interconnected and help explain the unique psychological hold that photographs have on people, which often leads to theorists ascribing ‘mystical’ qualities to the medium (Maynard 116). As I continue to argue, there is nothing especially ‘magical’ about photography, including ‘deepfakes,’ which I say are instances of photographic technology, understood broadly as a technology for marking surfaces for display.

In order to understand the first function, depiction, is it crucial to understand the notion of markings, as “markings put down intentionally on surfaces can and often do serve multiple functions” (Maynard 27). A painting cannot come into being without marking a surface with paint; similarly, a drawing is created from a pencil leaving graphite on a surface. It is not just the material used for marking, but also the surface itself, that form the depiction (Maynard 26).

Markings are often irreversible, but that is not always the case. ‘Deepfakes,’ and other digitally captured and displayed videos, are also forms of surface markings, “in the extended sense of being physically affected in a discontinuous way” (Maynard 53). Even Rodowick, who says “there are as many continuities as discontinuities” between analog and digital film, should agree that the changing colours of pixels constitute a marked surface, even if said markings are temporary and caused by coded signals (30). A marked surface need not be permanent to be considered as such, and a digital screen should be no different. I do not, however, agree with scholars such as Rodowick, who think this shift from analog to digital methods of display—from

a film-reel projector to a digital projector or an electronic screen—has wider ontological or epistemological implications, as an effect of the technical changes themselves.¹² Regardless of method of display, markings can be inadvertent or purposeful (Maynard 30). *Child in Backyard*, the cover photo of Maynard’s book, offers a few examples, allowing us to reflect on the types of markings we encounter all around us.



Fig. 2. Walker Evans. *Child in Backyard*. 1932, The Museum of Modern Art.

The 1932 Walker Evans photograph depicts a young, freckle-faced girl leaned up against a wooden fence and gazing off to her left. The photo itself is the result of photo-chemical markings

¹² I go into more detail as to why in Chapter 3.

onto a film surface. The photo is then transferred onto the page of the book by a printing process (Maynard 30). The depicted image itself shows many kinds of markings, including the etched drawing on the door, the scratches from the door latch, and the weathered wall behind the girl. Her face is also covered in freckles, which, like photography, are marks created when light strikes a surface.

The freckles themselves, though, unlike the photograph of the freckled girl, are not a kind of *display*, which Maynard succinctly defines, in this particular context, as “the function of making itself visible” (33). Freckles are not created from sun exposure in order to make themselves be seen; rather, their visibility is a by-product of their actual function, which is that they contain extra melanin to help block harmful ultraviolet rays (Sisson). It is incidental that they are visible, but they do not derive their purpose from *being* seen.¹³ A photograph, on the other hand, is a kind of marking that displays. The purpose of freckles is to be visible—not necessarily visible to humans, but something visible, nonetheless. To use photography, regardless of the specific process or intended use, is to create something that is to be seen, whether by humans or in transferring an image to another piece of technology, such as is the case with computer microprocessors.

The face of the photographed girl is not naturally freckled with the intention of being displayed, “although this opens the interesting topic of how accidental markings may be turned to display” (Maynard 30). These inadvertent markings, when presented in the photograph—itsself a display created by purposeful markings—allow for discourse. The freckles contribute to what we contemplate when looking at this photograph, for example: normative beauty standards,

¹³ Here, Maynard is making the distinction between natural and cultural signs. The argument in his book is not built on semiotics, which he rejects as “suggesting a mythical common function for all such markings” (60).

given the girl's overall tomboyish appearance; the hard life of a Great Depression-era child, whose face has been out in the sun out of necessity; or of her carefree childhood spent exploring the back alleys of her hometown, with the sun-kissed face to prove it. These examples show how we can use markings created without human intervention for discourse, just as much as we can to purposeful markings. We need not worry if the photographer intended to represent her freckles for a particular purpose, nor if he had intentionally drawn on each individual freckle; the intentional and unintentional, or what we call *automatic*, both contribute to our understandings of an image. I will return to the idea of automation later in this chapter, for it is strongly connected to the false claim that more automatic forms of images are either more or less truthful.

Depiction

For Maynard, one function that all forms of photography share is that they *can* depict.¹⁴ This should hardly come as a surprise; as Maynard explains, “the assumption that photography is essentially a depictive device and that its other uses are marginal” has long prevented us from properly understanding photography (24). Though this is a shared function across all photography, there are many other depictive technologies. Pencils, and paint and paintbrushes are also depictive technologies, which can be used to create depictions.

It is through photography's depictive qualities that people can imagine. Pictures, in particular, induce and shape our imagination; specifically, “marked surfaces...mandate that we imagine *seeing* things” (Maynard 104, emphasis in original). This does not mean that markings are invariably mimetic; rather, their very presence, even in the most rudimentary marks, can prompt us to imagine. The crudest line in chalk can resemble a snake to the imaginative mind,

¹⁴ That a technology can serve a certain function does not mean that it always will. Maynard uses the example of the microchip circuit, which uses silver-based photolithography but is incredibly small, to show that depiction need not be present (58).

regardless of the goals of whoever drew it. As just mentioned with the example of the photograph of the freckled girl, markings need not even be intentionally made to have discursive use. By the very definition of a marked surface, of which photographs are an example, “our act of looking at them is the act of looking at what they depict” (Maynard 104).

Depiction can be complicated at best, for there is the undeniably subjective experience in imagining. For our purposes, it will suffice to simply understand two facts: one, that depiction is a function shared across all types of photography (though not exclusive to photography); and two, that while depiction is a material characteristic of the medium, it is also deeply connected to our own subjective experiences of the world, and thus is simultaneously a part of, and separate from, the medium itself. This paradox, while intriguing, is outside the scope of this thesis, but it is worth mentioning here as an example of the complex interplay between medium and human. Depiction is an ontological quality of photography, but it is also, in part, a cognitive process, which allows us to imagine.

A persistent misunderstanding of photography is the conflation of imagining and depicting. As Maynard points out, there has been “a failure to develop a simple terminological distinction between *a photograph of something* and *a photographic depiction of something*” (114, emphasis in original). Maynard uses the film *King Kong* (Merian C. Cooper and Ernest B. Schoedsack, 1933) to illustrate this. There are at least two things to distinguish: a photograph of a small gorilla figurine, and a photographic depiction of the fictional creature King Kong (114). “A photograph of” refers to the subject that had been photographically captured; in the case of *King Kong*, it is the model gorilla that was animated, which is an unassuming toy. “A photographic depiction of,” on the other hand, incorporates the induced imagining that occurs from seeing the photograph. While a small gorilla figurine was filmed in *King Kong*, the film is a

photographic depiction of a giant, ferocious monster. Even if the audience is aware that *King Kong* was made by filming a small figurine, they imagine that it is depicting something else. There is this duality in every photographic image, regardless of its intended use. The photograph I took with my phone to remember where I parked my car is depictive, just as a stylized Man Ray portrait in a gallery is depictive. The difference between these photos is the use (for aiding memory versus for provoking thought). The use or uses, in turn, influence whether the viewer *focuses on* the depictive aspect—“a photographic depiction of”—or the detective aspect—“a photograph of.” It is important to remember that these two functions, depiction and detection, are not mutually exclusive; they rely on each other for making meaning, in addition to “background information regarding sources, as well as about our channels of perception” (Maynard 144).

Indeed, this distinction Maynard makes between “a photographic depiction of” and “a photograph of” succinctly describes the two functions shared across all photography: depiction, as I have described, and detection, which is the second one I will explain in more detail.

Detection

The other function of photography, according to Maynard, is detection. This is what seems to have caused the most extreme mystification of photography, which has led to an exaggerated degree of trust in photographic representations. I find that Maynard’s account of detection is a more straightforward way of characterizing the relationship between referent (that which is photographed) and the resulting image, rather than that of Bazin, whose writings have been foundational to film theory.

While even the earliest film theory pondered the philosophical role of the camera, André Bazin’s “The Ontology of the Photographic Image” is the most evocative and influential essay on the subject. It is also easily the most contentiously debated essay, for he touches on such lofty

topics in eight short pages. His early death at age 40 in 1958 further compounds the uncertainties of his theories. Regardless, different readings of this essay can support and refute claims of camera's fundamental distinction from other media because of its technological creation.¹⁵

Bazin states that the image “shares, by virtue of the very process of its becoming, the being of the model of which it is the reproduction; it *is* the model” (14, emphasis in original). This opaque description may seem as though he is equating referent and image, but it is more likely that the emphasis with italics of “is” indicates a less obvious connection. Suffice to say, this essay has been continuously debated since its publication, with no clear consensus on what he means, due in part to the often bafflingly vague and mystical language he employs elsewhere. At face value, Bazin's words in this essay describes light as having agency: “by the power of photography, the natural image of a world that we neither know nor can know, nature at last does more than imitate art: she imitates the artist” (15). This is where I find Bazin's argument the least credible, insofar as he is attributing the humanlike quality of agency to a nonhuman entity.

Another way scholars have tried to describe this referent-image relationship is by making direct comparisons between cameras and our mind, or our way of seeing the world. Hugo Münsterberg, writing in 1916, posited that the close-up “has objectified in our world of perception our mental act of attention,” in that it allows us to see the world without as little interference from our human subjectivity as possible (87). Cavell's work in the latter half of the twentieth century also draws on this desire to escape our perceptions. He argues that “the idea of

¹⁵ Daniel Morgan is one such scholar who gives a more generous interpretation of the relevancy of Bazin's work in the twenty-first century. Morgan argues that “his metaphors represent a series of attempts at understanding the peculiar ability of photographs to give us more than a representation, however direct and unmediated” (452). Furthermore, “Bazin seems to be saying that photography removes the object from any specific position in time. The objects a photograph presents may not exist in the present, but they are not exactly in the past, nor are they in any other time. They are real but outside (historical) time altogether” (452). I remain unconvinced, the reason being mainly that Maynard provides a more straightforward, logical account. See Morgan, Daniel. “Rethinking Bazin: Ontology and Realist Aesthetics.” *Critical Inquiry*, vol. 32, no. 3, 2006, pp. 443-481.

and wish for the world re-created in its own image was satisfied *at last* by cinema,” though he acknowledges that each of the arts is aiming for this fully objective realism (39, emphasis in original).

Maynard’s thinking on the relationship between referent and image is much clearer and relies less on abstract characterizations that give photography mythical qualities. After all, the mind and the search for objectivity do not teach us much about what cinema is, given how little we know about either (Carroll, “Film/Mind Analogies” 497). Maynard is not saying that photography, by nature of its ability to detect light, gives us unique insight into our own subjectivity in the world. Technology allowing us to do things we were previously unable to does not mean that we can better understand our minds. Rather, photography is yet another technology, such as pencils or other tools for marking surfaces, that extends our ability to do things. Photography allows us to create enduring images using light, which we are unable to do on our own.

One example of photography that emphasizes the detective qualities is the use of photo-finishes in sports. This is another specific purpose of photography, one that confounds typical accounts of photography that are prescriptive in its uses and, thus, its ontology. In his book on the phenomenon, Jonathan M. Finn says:

It is sufficient to note that photography’s adoption into various social and institutional practices and its status as a purveyor of truth should be seen as a product of both photography’s nature and its culture. In the former, the photomechanical apparatus was and is able to produce highly verisimilar images in ways and quantities that other modes of visual representation (printing, drawing, painting) cannot. In the latter, the gradual accumulation of a discourse of objectivity surrounded the photograph as it was

increasingly employed across a range of social institutions and practices from the late nineteenth century. (11)

I disagree with Finn's assessment that photography's ability to quickly produce highly detailed images means it can be used more truthfully than any other medium, even in this particular instance. In fact, this forms the basis of my thesis, as it directly relates to the misconception that, by contrast, 'deepfakes' are an inherently deceptive medium because of what the technology does to the supposedly inherently truthful medium of photography. That a camera can detect and depict light on a marked surface after shutter exposure does not indicate truthfulness. As I will discuss in Chapter 2, this is a too simplistic characterization of truth. In terms of Finn's assertion that photographic truth is a product of culture, I do agree that we treat the results of photo-finishes, for example, as truthful because of a "gradual accumulation of a discourse of objectivity." Unlike Finn, though, I see this as, at best, an oversimplification of truth and, at worst, a glaring flaw in our critical accounts of photographic images.

Likewise, Maynard also discusses the photo-finish in his book. He argues that they are a good example of how detection and depiction rely on each other to produce meaning:

It was the operator's setting of the film's scanning speed that made the runners' optical smears look anything like a *picture* to us in the first place. Film-rolling rate for a horse race might be five centimetres per second, in a mechanism adjustable between one-twelfth and four times that speed. (138, emphasis in original)

This account of preparing a camera for a photo-finish demonstrates the amount of human intervention required just to take a photograph capable of depiction. Despite the work involved in taking a photograph, there is still the common attitude that humans play a secondary role to the supposed automation of the camera.

Automation and Intentionality

Like photography, popular accounts of ‘deepfakes’ have also emphasized their automated nature. This has led to fearful attitudes towards such technology. However, we need to clarify what it means to be automated. As the above example of taking a photo-finish shows, photography is not as automatic as we are often led to believe.

All manner of tasks have been automated by technology, but photography best exemplifies the introduction of automation into image making. Historian Robert D. Friedel notes that the advent of photography more clearly melded the arts and sciences than any other invention during the Industrial Revolution (341). The technology was strongly associated with “science and the growing sense that nature’s secrets were truly at hand for the creation of new devices and powers,” which was a prevailing cultural attitude of the mid-1800s (341). Early accounts of photography emphasized its scientific spectacle, as well as the possibilities for artists who, at that point, would use the *camera obscura*¹⁶ to trace images.

Likewise, Mary Warner Marien notes that photography “was the only one of the early-nineteenth-century transformational technologies that forcefully connoted the lack of control by human beings” (169). The ability to make such precise markings on a surface with minimal human intervention—just prepare your camera, press the button, and, depending on the type of photography, develop your pictures—has elicited strong reactions from both sides. Painter Paul Delaroche, upon seeing an early daguerreotype in 1839, is quoted as saying, “from today, painting is dead!” (Friedel 341; Maynard 54) (As Maynard points out, he said this when now-

¹⁶ Per the Oxford English Dictionary: “An instrument comprising a darkened room or box with a convex lens or a pinhole in one side, used for projecting an image of an external object on to a surface inside the instrument so that it can be viewed, drawn, or (in later use) reproduced on a light-sensitive surface” (“Camera Obscura, n.”).

famed painter Paul Cézanne was an infant, disproving the imminent mortality of painting at the hands of photography (54.)

Bazin's work also characterizes film as innately automatic, which he contrasts with other, older media: "all the arts are based on the presence of man, only photography derives an advantage from his absence" (13). He says this absence is what allows the artist to focus on self-expression: the photographer's inherent subjectivity is brought into the photography by way of selecting what to photograph and "by way of the purpose he has in mind" (13). According to him, the artist using photography no longer has to worry about being able to create perceptually realistic art, which allows art to "recover its aesthetic autonomy" (16).

Automation discourse continues into the twenty-first century. While digital photography remains in the throes of debate with regards to the ease of taking a photograph, the rise of computing has added further challenges. The 'deepfake' is one such technology in these crosshairs. However, 'deepfakes' are no different from photography, in that they are not wholly automatic: even the most sophisticated algorithms will need some manual corrections, and making the algorithm in the first place requires humans to design and test the underlying code. Like all amplifiers (and suppressors) in technology, automatism is a characteristic that can be seen as positive or detrimental.

Medium

Technology and medium are often discussed in conjunction, since technologies (photography) can be used to produce a medium (photograph). While the meaning of "medium" may seem obvious, 'deepfakes' confound our everyday understanding of this concept. By definition, 'deepfakes' necessarily use a combination of photographically captured video and computer-generated imagery. This leads to the question of whether or not 'deepfakes' are a new

medium or a continuation of the moving-image evolution. Is the ‘deepfake’ video one whole medium, or is it a second medium, with one medium (the AI face) placed onto another (photographic video)? And is there something about ‘deepfakes’ that makes them ontologically distinct from all other media, especially other moving images?

Guillory’s historical account of the term’s use reveals a complicated relationship, one that provides insight into Western philosophy’s changing approaches to the study of communication. It was not until the advent of such nineteenth-century inventions as telegraphs and phonographs that the concept of media was raised (321). Similar to contemporary discourse surrounding digital media, these new technologies of the era led to critical re-examinations of all preceding media.

It was only once these communicative technologies emerged that philosophers developed the concepts of medium and communication to aid discourse on art and representation, where analysis was previously dominated by mimetic inquiry (Guillory 322). Though *Poetics* greatly influenced Western thought on imitation or mimesis, here Aristotle creates abstract divisions between paintings, songs, dance and poetry, based on how each category should best use mimesis (Guillory 322-323). For Guillory, this is an allusion to the concept of what would become the medium, where translators lack adequate vocabulary to describe what Aristotle ultimately describes as “in different things” (323).

Broadly speaking, there are three ways to conceptualize the medium: in its material condition, in its creation, and in its being experienced. To be clear, there can be a plurality of definitions of medium. In some contexts, it is useful to talk about the material composition, for instance. I am not attacking these definitions; rather, I want to interrogate the implications of these terms when used in the metaphysical, epistemological context in which my thesis is

engaged. Taking a closer look at each definition will reveal the interconnectedness in our understanding of technology.

Materials

The material is one such defining characteristic of a medium: the canvas and oils of a painting, or the marble of a sculpture. This definition, however, soon gets complicated when discussing digital creations. While traditional celluloid has a tangible materiality, what about digitally created and stored video? Is its materiality defined by the hardware it is stored and processed on, the pixels that are alight when the code is read, or some combination thereof? There is also an almost infinite number of potential media. Since a sculpture can be made of stone, clay, metal, glass, plaster, or any number of other materials, is it really of much use to ascribe such narrow definitions to a medium based on its physical components? In Noël Carroll's polemic essay "Forget the Medium!", he points out that "artforms possess more than one medium, and many of these media themselves have divergent and nonconverging potentials" (6). In short, defining a medium by its material overly broadens and restricts categorization at the same time.

Creation

Another way to define the medium is to focus on its method of creation. You can distinguish a painting from another medium because you know that the painter used a paintbrush to apply the paint to the canvas, for example. However, it is still too difficult to pigeonhole creations into these uses. Take, for instance, camera-free films, such as Len Lye's *A Colour Box* (1935), Norman McLaren and Evelyn Lambart's *Begone Dull Care* (1949) or Stan Brakhage's *Mothlight* (1963). These works were created by painting on or pasting objects directly onto the celluloid, which was then projected like any photographed strip. This is not the 'intended use'

(whoever decides that) of the film medium, as these works do not use the photochemical possibilities of celluloid coated with photographic emulsion. You could classify these examples as the distinct medium of animated film, as opposed to photographically captured film, but you will soon find this, too, can be overcomplicated, especially when you use materials as a criterion for classification.

This raises another question, though: how *do* you know that the painter used a paintbrush? What if the artist actually used their fingers, but the effect is indiscernible from paintbrush strokes? If you were to define the medium by the act of its creation, you would have to separate paintbrush paintings from finger paintings, and so on with any possible tool for applying paint to a surface. This information is also largely unavailable to us, as we are not capable of knowing what tools were used for every depiction. The very basis of the *First Draft* articles hinge upon the fact we are not always aware of when AI is used to make a depiction.

Experience

A third option for defining the medium is by experience. I know a sculpture is a sculpture by looking at or touching it, and I know that a film is a film from seeing and hearing it. A sculpture can be any of the aforementioned materials, and a film can be projected on celluloid, lit up on a screen, or any other possibility yet to be created. Such a simple definition allows for flexibility in materials and methods of creation, as well as potentials for mixing of media. This also aligns with the new technology of the ‘deepfake,’ which, as described, is made using two different techniques (photographic capture and computer generation).

It may look as though this is the preferred way to define “medium,” at least compared to the previous suggestions of material and creation. Central to this thesis, the ‘deepfake’ appears to confound these other boundaries of a medium, since its creation is in part via photographic

capture (the target video) and AI digital synthesis (the face). How do you reconcile the two methods involved in making a ‘deepfake’? It is rather straightforward: while only the face of a photographically captured video is synthesized using computer generation, the entire video is recontextualized into a ‘deepfake.’ Much like Maynard’s emphasis that “*unmarked* parts of surfaces make up parts of images” (26, emphasis in original), we should also define the ‘deepfake’ as including the non-synthesized parts of the video. Thus, it would be fair to refer to a ‘deepfake’ as a singular medium, rather than a pastiche of two media, at least as far as the philosophical concept of medium is concerned.

However, it is not simple enough to dispense with all notions of material and production, in favour of a wholly experiential definition of the medium. We cannot completely disregard the method of creation when defining “medium.” While all media have emotive possibilities, photographic images and videos are the predominant visual tools for journalism, in addition to the written word and its ability for us to imagine. It is possible to make highly effective drawings or paintings to convey the brutal reality of war; however, the unique affordances of photography account for its enduring popularity in reportage. This is where Maynard’s theory of technology comes in. The depictive and detective functions of photography can emphasize one function over another in our use of a photograph. This does not mean that photography is more truthful than drawings or paintings, by virtue of its detective function. Rather, it is a combination of detection and certain levels of automation within photography that makes it a particularly effective technology for journalism. It is by no means the ‘best’ or ‘optimal’ kind of photograph, since these are contingent upon human values for using photography.

Essentialist Accounts of Photography

Not all amplifiers will be beneficial or desired in all scenarios, however. The overarching theme of Friedel's book *A Culture of Improvement: Technology and the Western Millennium* is his conviction that "the history of technology is in large part the history of achievements measured by such an implicit standard—what has sometimes been referred to as 'progress talk'—[which] is one of the primary hurdles that historians have been struggling to overcome" (3). Though he denies any philosophical allegiance in the introduction, this account of technology as history without predetermination counters a Hegelian vision of continuous improvement (Friedel 10). Friedel's description of keystone technological inventions demonstrates the uneven path Western civilization has taken over the last two millennia. While all emerging technology builds upon past discoveries and inventions—such as the use of photography in computer microprocessors—there is no absolute value we can ascribe to any technology (Maynard 6).

As discussed earlier in this chapter, the concept of progress is highly contingent on human values, which are subjective and malleable. What is largely considered progress may eventually be deemed a setback, either by an individual or as a collective. For example, when plastics were invented, we appreciated the cheap manufacturing cost, the lightweight material, and other properties that led to its mass adoption. Of course, we have since seen the negative effects associated with the materials, both to the environment and to human health, and are in an active process of creating alternatives. Since there is no universally agreed upon metric for improvement, we cannot say that technologies have been getting better overall in any general or absolute sense. A recurring theme throughout this thesis, as is the case here, is that these value statements are overly broad. Only when we make a specific claim—for example, that current

digital photography is superior to drawing *for quickly making images*—can we have productive conversations around the subjective values of different media.

Despite this, the Hegelian view of technology as a series of linear improvements (or decline) is widespread in the discourse on photography and photographic media like film, with many of the formative scholars seeing photography as a radical departure from previous methods of depiction. We can call these thinkers *essentialists*, for they see photographs and other media as having a distinct essence, or ontology, which should inform how the medium is used.

Siegfried Kracauer is one such essentialist whose work, in his own words, “rests upon the assumption that each medium has a specific nature which invites certain kinds of communications while obstructing others” (3). He follows the positivist line of thinking, whereby proponents “[aspire] to a faithful, completely impersonal rendering of reality” (Kracauer 5). This is in line with popular Western sentiment of the nineteenth century, when science had emerged as the major force for explaining *all* phenomena, art and philosophy included. Though Kracauer denies fully adhering to positivism, referring to its proponents as “[lending] support to the cult of realism, thereby proving detrimental to the elevated art,” he nonetheless shares their preoccupation with what he sees as photography’s innate capacity for realism (7). Kracauer takes a somewhat more measured approach to differentiating photography from other media, most notably from paintings. He argues that “if any medium has its legitimate place at the pole opposite that of painting, it is photography” (13). For him, the main aesthetic error photographers have been making, at least in terms of artistry, is to make photographs too painterly; when this happens, the photograph has veered too far to painting on this axis defining the two media. In order to successfully work with photography’s affinities, Kracauer says there needs to be an element of spontaneity coupled with artistic contemplation, or “the desire to

externalize his inner images and the desire to render outer shapes” (17). I argue that he is making a false equivalency between artistry and the direct creation of the artwork; designating an object as art after it has been made is just as powerful a method, as Dadaists like Marcel Duchamp demonstrated in the early twentieth century. Artistry does not only come into being at the moment the photographer takes the photo, but also when it has been developed and examined. To summarize, Kracauer’s essentialist account relies on the misconception that the medium or technology employed has a direct bearing on its uses. The technology or medium does not dictate how best to make a representation, for the criteria of what qualifies as ‘best’ is highly subjective and changes depending on the context.

Summary

Returning to the two-part series on labelling AI use, I hope I have succeeded in convincing you that such a regimen would be foolish—and possibly even dangerous. Apart from questions of practicality and implementation, these labels perpetuate the false narrative that deception is inherent to the technology or the medium. These labels also ignore the true source of truth and falsity: human intent. As I will describe in the following chapter, it is possible to lie using any medium, as much as it is possible to be truthful.

Nonetheless, it was critical that I consider the persistent argument that the medium and technology have a direct bearing on the truthfulness of a representation. This is particularly crucial given that my subject of choice, the ‘deepfake,’ has a name that erroneously ascribes negative qualities. It should also be clear by now that technology and medium are deeply enmeshed, for the technology’s amplifiers influence the ways in which an image can be created. The difficulty in creating discrete boundaries from one medium to another demonstrates a futility, insofar as ontology is concerned, in ascribing distinct qualities to one medium over

another. With regards to their truthfulness or lack thereof, this is a fruitless argument: all media are equally capable of being used for truth or deception. Saying that ‘deepfakes’ are ultimately photographs, or computer graphics, or a combination therein, has no bearing on their capacity to be used for telling the truth or for fabricating lies. Classifying ‘deepfakes’ into a specific medium would not change the most salient aspects of truth: that *all* media are equally useful for deception. It is also worth emphasizing the impurity of boundaries between any media: writing on a painting, music in a play, and the painted backdrop in a film are all examples of melding media into a singular medium.

Nonetheless, it is worth determining what kind of medium the ‘deepfake’ is, insofar as it will allow for further analysis of this relatively new phenomenon. I argue that the ‘deepfake’ is, in fact, a photographic medium. ‘Deepfakes’ are marked surfaces made “through the agency of light and similar radiations” (Maynard 3). The addition of computer synthesis does not change the fact that the ‘deepfake’ face bears a strong causal relation to the photographically captured face. For one, this synthesized face is created from the algorithm analyzing a massive body of photographically captured images and videos. Secondly, the algorithm creates a new face that is still linked to the detective qualities of the source video using facial mapping.¹⁷

Having established that no medium is inherently more truthful or deceitful than another, I will continue this argument into Chapter 2. Here, I will discuss truth, falsehoods, lies and reality more generally, to demonstrate how such value judgments are human-made conceptions.

¹⁷ I describe facial mapping and similar technologies in Chapter 3.

Chapter 2: Characterizing Truth as a Human-Made Action

If truthfulness and falseness are not inherent material characteristics, you may be asking, where, then, do we locate these values? Having a critical understanding of *what* truth and reality are will further solidify my argument that such values are not dependent on the medium or technology. What is truth? And what does it mean to lie, or to unintentionally make a false statement? These are profound moral questions. While I cannot fully address these complex, centuries-old problems, I can provide a sound basis in which to situate my particular example of ‘deepfakes.’

As I will argue, it is not a simple dichotomy whereby truth and reality, and representation and falseness, are directly correlated. Rather, these four terms are interwoven. Further complicating these broad epistemological concepts is the addition of intent when discussing falsehoods. What does it mean to lie, or for something to be fake? I posit that there is an intention to lying, making it a uniquely human action, and, therefore, only attributable to the contextualization of representation. The very moniker, that of a ‘deepfake,’ denotes a kind of deception, supposedly characteristic of the technology, understood as an alteration of an originally and essentially photographic image and thereby a diminution of a supposedly photographic veracity. As I intend to make clear (and as demonstrated through use of my scare quotes around ‘deepfakes’), this technology is not inherently a lie. Likewise, ‘deepfakes,’ or any other depiction, are not inherently truthful.

In this chapter, I will begin by defining truth, as its multifaceted nature is often oversimplified. Specifically, I want to draw our attention to different ways of conceptualizing truth, one of which is more productive for our purposes of truth as a discursive action. I will then

demonstrate the ways in which we use this particular form of truth for determining not only truthfulness, but also if someone is lying.

Understanding What is (and is Not) a Type of Truth

How do you define truth? This is one of the oldest philosophical questions, and there is no easy, universally agreed upon answer. Plato's famous cave allegory helps illustrate this complexity. In book ten of *The Republic*, his treatise on the ideal society and its citizens, he describes a hypothetical situation in which people are imprisoned in a cave, unable to move their head and only able to see shadows illuminated by a fire. It is only once they are unbound and able to leave the cave that they have access to truth:

The visible realm should be likened to the prison dwelling, and the light of the fire inside it to the power of the sun. And if you interpret the upward journey of the soul to the intelligible realm, you'll grasp what I hope to convey, since that is what you wanted to hear about. Whether it's true or not, only the god knows. But this is how I see it: In the knowable realm, the form of the good is the last thing to be seen, and it is reached only with difficulty. Once one has seen it, however, one must conclude that it is the cause of all that is correct and beautiful in anything, that it produces both light and its source in the visible realm, and that in the intelligible realm it controls and provides truth and understanding, so that anyone who is to act sensibly in private or public must see it.

(Plato 517b-c)

Plato uses this allegory, in part, to demonstrate the fallibility of our own perception, specifically of sight. As mentioned in the previous chapter, Plato has a skeptical philosophy, wherein we are the metaphorical prisoners, falsely believing we know the nature of reality, and thus truth,

because of the ‘shadows’ we see. It is only in emerging from this allegorical cave that Plato believes our souls will be cognizant of the truth.

I do not agree with this skeptical view that Plato holds. For one, our experiences are inextricably connected to our perceptions of the world. It is only through these subjective senses that we can access the world. If our only access was through our subjectivity, how would we even know if we have transcended into an objective experience of the world?¹⁸ There is no need to dwell on this “intelligible realm” if it is impossible to ever access, and thus even determine its existence. I want to shift our focus away from this conceptual impossibility, towards a more concrete system of knowing. As philosophers before me have posited, truth can be thought of in two categories. Here, I will use the distinctions of “Truth,” capitalized, and “truth,” all lowercase. These two types (Truth and truth) are analogous to Plato’s distinction in the cave allegory.

Truth as Objective Reality

What Plato describes as the “intelligible realm,” which encompasses the nature of our world in its objective entirety, is what I will call Truth, with a capital ‘t.’ This is immutable, omnipresent and not bound to any person’s sensorial experience of the world. Since we are all restricted to our own perceptions of the world, Truth is a wholly theoretical concept, impossible for any individual, or even a collective of people, to grasp in its entirety. Nonetheless, Plato seems to think it is possible to achieve an all-knowing access to this Truth, which is what we do when we search for any truth.

¹⁸ W.V. Quine, who could be described as anti-Platonist, also makes this point. He begins *Word and Object* by saying that his desk “manifests its presence” when he touches and sees it. Moreover, he insists that “physical things generally, however remote, become known to us only through the effects which they help to induce at our sensory surfaces” (1). See Quine, W.V. *Word and Object*, new edition. MIT Press, 2013 [1960].

Humans have a deep-seated anxiety in not knowing the Truth. What is so difficult when discussing truth is that overthinking seems to be inevitable; that we can never possibly know *everything* can lead to a defeatist attitude. A common scapegoat when discussing truth in any form is to invoke this capital ‘t’ Truth, with such refrains as “we could never possibly know the truth for sure, so why bother trying?” To reach this conclusion is to confound these two categories. The criteria for determining a specific truth do not need access to capital ‘t’ Truth. That we lack access to all Truth does not diminish from our ability to collectively assign truth to specific claims. After all, what advantage is there to be gained from having access to Truth? As Michael P. Lynch says, “Truth is always truth in context; what is true for one person might not be true for another” (32). Such relativism is not to be feared, though, he argues. “Relativism is not necessarily a threat to truth’s value. The common temptation to think otherwise is the result of another myth: that truth matters only if it is absolute” (32). That is, I do not need access to an additional realm of understanding, to transcend the inherent subjectivity (or relativity) through which I navigate the world, in order to understand that what I see, hear, feel, smell and taste is an apple.

Truth as Subjective Experience

As mentioned in the previous chapter, I disagree with Plato’s skeptic philosophy. It is impossible for anyone to have access to all Truth of the world, but that should not distract us from seeking out lowercase ‘t’ truth. The “knowable realm” of the cave allegory, the lowercase ‘t’ truth, is that which we know to be true through our mental processes. In her influential 1978 book, *Lying: Moral Choice in Public and Private Life*, ethicist Sissela Bok contends that the impossibility of grasping the whole truth “has very little to do with our choices about whether to lie or to speak honestly, about what to say and what to hold back” (4). What we see, hear, smell,

taste and touch inform us of what we deem to be true. I know this is an apple, for example, because I see its red, round shape, I hear and feel the crunch, and I smell and taste the apple. We are constantly creating new truths as we go about our lives in this world, whether we want to or not. The defining characteristic differentiating Truth from truth is the intervention of human agency: while Truth is defined by absolute objectivity, truth is inextricably subjective.

Optical Illusions

As a counter to my argument, consider the *trompe l'œil*, or optical illusion. These are representations that, when viewed in certain contexts, ‘trick’ the brain into seeing something that is not real. A common example is sidewalk chalk drawing of a sinkhole. The angle and attention to detail can lead a passerby to think, however briefly, there is a sinkhole in the concrete. An optical illusion can also be achieved all on its own, without the illusion relying on its placement in relation to the real world, as is the case of the duck-rabbit image, which can resemble a duck, a rabbit, or both simultaneously. Optical illusions are also naturally occurring; Plato refers to a straight object appearing bent when partially emerged in water, which he uses to conclude that “every other similar sort of confusion is clearly present in our soul” (Plato 602d). I bring up this example to draw our attention to the pervasiveness of this argument, that we cannot trust even the lowercase ‘t’ truth formed from our perceptions.

It could be argued that optical illusions are inherently deceptive, with which I disagree. Any visual item, whether it is a drawing or a mirage, is not in itself ‘tricking’ us because that would require sentience. As explained in the previous chapter, truth is not a material quality of any medium, and this applies also to deception. The elements that could distinguish an optical illusion from any other representation—technique of its construction and placement in relation to the real world—are deliberate choices by whoever created the illusion. For naturally occurring

illusions, they do not deliberately ‘trick’ us at all. The modern thinker, equipped with a basic scientific understanding, should be accepting of the fact that light can refract in ways that appear odd to our eyes. The reason this happens is not for the purpose of deception, much like how, in Chapter 1, I argue that freckles are not created to be seen. That a scientific phenomenon occurs does not mean that what we see is deliberate; in the examples of Plato’s stick in the water, or with freckles, the visual effects are incidental. To try to deceive, just like to tell the truth, is an action that requires sentience.

The above examples of optical illusions are all based on our sight. However, there are other senses we normally use for accessing the lowercase ‘t’ truth, that can achieve illusory effects. Notably, there are cinematic gimmicks, such as 4D (four-dimensional) movies, where, in addition to sound and hearing, there is haptic feedback: viewers could see an airplane take off, hear its engines roar, and also feel their seats vibrate and move accordingly. This type of entertainment, commonly used in movie theatres and amusement parks alike, is marketed under the assumption that people want to experience greater realism. While realism denotes a resemblance to reality, this is not the same as experiencing truth, despite our perceptions of reality shaping lowercase ‘t’ truth. As I will describe later in this chapter, when I present the case study of a famous Roger Fenton photograph, our perceptions alone do not inherently make compelling truth claims. This is because truth, reality and perceptual realism are separate concepts that interact with each other.

Perceptual Realism

As discussed in Chapter 1, the medium of a representation does not have its own affinities. Similarly, the medium does not determine how realistic the art is. In order to discuss the concerns surrounding ‘deepfakes,’ it is imperative to question our relationship to realism. To

be concerned with ‘deepfakes’ is to assume that it is immoral for something to be *too* realistic. This leads to a host of questions: Is there a point at which a representation is too realistic, and how would one define such a point?

One of the most apparent worries that arises from AI-generated video faces is its high degree of perceptual realism. That is to say, these videos aim to mimic the ways in which we experience the world through our senses. The desire for greater perceptual realism is what guides further developments of this technology: better visual cohesion between AI and non-AI components, more humanlike blinks and other natural tics. There is a persistent fear in media narratives of ‘deepfakes’ that, as technology becomes more sophisticated, the distinction between representation and reality will irreversibly blur, or that we will no longer be able to make the distinction.

Fiction and Non-fiction

It is in fiction where we tend to be the most willing to accept the complicated tension between true and false. When we watch a film, for instance, we accept that the story being presented is a fictional construction. Nonetheless, even the most esoteric fictional stories have a certain amount of grounding in the real life; a story that is completely devoid of any relatability to human experiences would be incomprehensible. *Star Wars* (George Lucas, 1977), for example, may be set in faraway planets with imagined spaceships and aliens, but the characters and the plots must bear some resemblance to our own experiences. Even if the idea of blowing up the Death Star and overthrowing the Sith Empire is unrealistic, there is still the realistic story of political rebellion and the fight against imperialism and fascism. These are what Iris Murdoch refers to as “moral pleasures,” which she sees as “the highest pleasures of literature and, one might say, of art generally” (257). Murdoch is saying that literature is a strong vehicle for

conveying these messages of universal human experience. That does not mean she is advocating for only consuming literature that demonstrates positive ethics; rather, she is stressing the importance of personal reflection in reading, even if that is through immoral characters and situations.

What is it about literature—or film, in our case—that allows for this? Murdoch suggests that “art is very much to do with accident, with contingency, with detail, with self-expression, *with trickery of all kinds, with magic*. I think the idea of magic, here, perhaps sums up the extraordinary ambiguity of art, and of this Plato was, of course, very well aware” (247, emphasis is my own). Her use of the word “trickery” is more subtle than simply saying that art is lying to you. This trickery is all within the mind of whoever engages with art, not with the intentions of its creator. When interacting with art, you are persuaded into considering what Murdoch deems the “moral pleasures” of art (257). We are taken in by the aesthetic pleasures of engaging with art, from which we can then receive such “moral pleasures.” Storytelling is one type of art that can be used in this manner to great effect. It is why the Aesop fable of the ants and the grasshopper is more compelling to children than simply commanding them to work hard to plan for the future. This technique of storytelling for moral purposes has been used across cultures for centuries, in part because of its effectiveness.

The “moral pleasures” Murdoch describes need not be as obvious as my Aesop example. In engaging with art, we are able to relate to our own experiences of the world. Like Maynard, she emphasizes the importance of imagining, though for her it is “the ability to see the other thing, what one might call, to use those old-fashioned words, nature, reality, the world” (255). Fictional art and stories, despite their tenuous connection to reality, nonetheless contain

lowercase ‘t’ truths (since it is inextricably subjective), by the virtue of our shared connection to reality.

Using Truth Claims

When we are looking for the truth, far more often than not, we are looking for a very specific truth. For example, who broke this window? In this search for truth, we need not be concerned with Truth at large, nor with the metaphysics of windows as a whole. Rather, we are looking for a very particular type of truth, which can be determined with evidence accessible to our perceptions: the side of the wall in which the glass fell, the size of breakage, any objects on or around the area of breakage, and so on. In all of these criteria, we are judging whether or not this contributes to the particular truth you are seeking.

Other People’s Perceptions

In addition to our own perceptions, we can choose to accept others’ characterizations of their own perceptions (for example, “I saw a baseball smash through the window”). This lowercase ‘t’ truth is also what others inform us about, leading us to accept or reject information as true or false. There is a great deal of trust in the truthfulness of others: that the stories your friends tell really happened, that the news is accurate, even that the mechanic properly fixed the brakes on your car. We necessarily rely on others to be truthful to their own perceptive experiences, precisely because the omniscient Truth is impossible to access, and because our social workings are so complex that we cannot functionally participate without any truth of others. There may be discrepancies between the evidence we perceive and witness statements, and we can either accept one of the truth claims based on sufficiency of details, or we can offer a

less precise truth claim (such as “I think a baseball smashed through this window, but I do not have the ball to prove beyond a reasonable doubt it was, nor can I say for certain who did it”).¹⁹

To those who are still skeptical that we need not be concerned with the capital ‘t’ Truth in these inquiries, who say that we still do not *know for certain* the truth, I counter: these claims-specific truths are informed by subjective experiences of the world, which, yes, are imperfect. However, it is all we have access to, thus being far more productive than any esoteric notion of Truth, to which we inherently have no access. If new evidence were presented that adds more information, or that contradicts previous evidence, then we can re-examine the truth we have claimed. No truth derived from truth claims can ever be proven as the absolute truth, since Truth requires an absence of human subjectivity. A skeptic would see this as grounds to dispense with all notions of lowercase ‘t’ truth; I counter that there are varying levels of truthfulness to these claims that should not be ignored. While we can never prove any truth claim to absolute certainty, some truth claims have so much evidence pointing to their veracity that we can treat these claims with near-absolute certainty. If we receive further information to disprove the previously held truth claim, we can adjust; understanding the difference between truth and Truth allows for flexibility in reassessing lowercase ‘t’ truth claims.

It is on this basis that we should analyze ‘deepfakes.’ The technology itself does not mean all uses are inherently false, for we can still derive truth from them. This can be factual truth, as exemplified by the speech President Nixon gives in the MIT ‘deepfake,’ or part of the “moral pleasures” Murdoch describes.

¹⁹ As Ludwig Wittgenstein notes, describing the complexity of the social contexts within which we endeavour to establish certainty: “I learned an enormous amount and accepted it on human authority, and then I found some things confirmed or disconfirmed by my own experience” (23e). See Wittgenstein, Ludwig. *On Certainty*. Edited by G.E.M. Anscombe and G.H. von Wright. Harper and Row, 1969.

It is important to note that the search for truth is a critical undertaking in any society, as is having a generally agreed upon notion of truth. We must acknowledge the ambiguities of understanding truth, while, at the same time, not allowing this ambiguity to be weaponized into disinformation, mistrust and mayhem.

Evaluating Truth Claims

Now that we have established that truth claims are human-made actions, which can be evaluated by the lowercase ‘t’ truth, what does this mean for evaluating an image? As explained in the first chapter, images themselves do not contain truths; likewise, they are not inherently false. Despite this, we tend to treat images, especially photographs, as bearing the truth, and the use of staging or other alterations to the *mise-en-scène* is considered deceptive. This is seen in the account of filmmaker Errol Morris’s search for truth in a famous photograph.

In a three-part blog for *The New York Times* (later a chapter of his book *Believing is Seeing*), Morris describes his near-obsessive quest for the truth of a photograph that Susan Sontag had famously claimed was staged. The photograph in question, *Valley of the Shadow of Death*, was taken by British photographer Roger Fenton during the Crimean War in 1855. This photograph is one of a series taken during the war, including another photograph taken at the same angle as the famous shot. Sontag asserted that “the cannonballs are thick on the ground to the left of the road, but before taking the second picture – the one that IS always reproduced – he oversaw the scattering of the cannonballs on the road itself” (qtd in Morris 4). This offhand comment led Morris to travel to the shooting location in Ukraine, trying in vain to determine if Fenton really did move the cannonballs. He eventually concluded that yes, the photograph without the cannonballs on the road was taken first, because the movement of rubble indicates that they were initially in the ditches.



Fig. 3. Roger Fenton. *Valley of the Shadow of Death*. 1855, The J. Paul Getty Museum.

Even though Morris was able to demonstrate beyond a reasonable doubt that Fenton moved the cannonballs onto the road to take his second, more famous photo, does that make the photograph itself a lie? As previously stated, we use truth as a discursive action, without which there is no relevant truth; therefore, the photograph, on its own, is neither truthful nor deceitful to us. But is Fenton's use of this photograph untruthful? It is not always obvious in what ways someone is using a photograph—and thus the truth claim they are making—so we have to intuit from the context. We can guess, based on the whole series of Crimean War photographs, that Fenton wanted to convey the human cost of war; many of the other photographs from his time documenting the war include military officers, zouaves and Crimean Tatars in traditional outfits (“Fenton Crimean War Photographs”). If Fenton were to say that *Valley of the Shadow of Death*

shows the gravity of war in the mid-nineteenth century, then we could evaluate this claim as truthful. This is an effective photograph for imagining the many cannonballs raining down on this desolate country road. If, on the other hand, he was to say that the photograph is what this particular road, at this particular time, looked like when he stumbled upon it, then that would be false. Morris's analysis of the two photographs, coupled with knowledge of the order in which they were taken, disproves this. Based on the composition and subjects of the other photographs in his Crimean-War series, the truth claims Fenton is making are more likely to be the former.

Comparing the two photographs, with and without the cannonballs on the road, I would argue that the second one, with the balls on the road, does a better job of making me consider how it would have felt to be on that road in the midst of warfare. The first photograph, wherein most of the cannonballs are in the ditch, more so emphasizes the probability that these balls rolled into the ditches after hitting near and on the road. With that being said, it is worth pointing out that this truth claim—that *Valley of the Shadow of Death* shows the gravity of war—is far more compelling than saying that it shows the literal gravity of the cannonballs. As mentioned in Chapter 1, photography's detective function has, in part, led to us misattributing it as an inherently truthful medium. The truth that we use in our everyday lives is the lowercase 't' truth, which is based in claims made by people. Fenton, with his famous photograph, is emphasizing the depictive function of photography, but we are nonetheless moved by its detective function as well.

We can undergo the same critical evaluation of other kinds of representations, including 'deepfakes.' One such example features Morgan Freeman, Kim Kardashian, Donald Trump, Mark Zuckerberg, Marcel Duchamp and Freddie Mercury. These massively influential figures are all featured in a video praising Spectre, a big data company whose own influence seems

eerily close to that of Big Brother from *Nineteen Eighty-Four*. The nearly five-minute-long video is made up of ten short clips of each of the above celebrities alone in turn, speaking directly to the camera (Kardashian and Zuckerberg are each in two clips, while Trump is in three). These short monologues are separated by black backgrounds with stark white text, with such short, pithy titles as: “Friendly Totalitarianism”; “Digital Dadaism (Dataism)” —aptly describing the Duchamp clip—and “Hypernormalisation” (Posters, *Veridical Fakes*). The talking heads describe revering data as religion, the commodification of your privacy, data as an important currency that people willingly relinquish, and the ubiquity of such practices.





Fig. 4-5. Bill Posters and Daniel Howe. Stills from *Veridical Fakes*. 2020, billposters.ch.

Once again, reader, much like in the introduction, I am sure you suspected that these are ‘deepfakes,’ that the celebrities were not filmed making such brazen statements. Though the name is inspired by the infamous data scientist from the 2018 Facebook-Cambridge Analytica scandal, the Spectre here is a wholly fictional company (Posters, “The Art of Interrogation”). This video, oxymoronically titled *Veridical Fakes* (“veridical” meaning truthful), is what self-professed artists and hackers Bill Posters and Daniel Howe call “synthetic art” (Posters, *Veridical Fakes*). The clips from *Veridical Fakes* are part of a bigger project, which Posters dubs the “Big Dada” series from the Spectre project (@bill_posters_uk). Spectre, in turn, is also an immersive installation that premiered in 2019 at Sheffield DocFest (Posters, *Veridical Fakes*).

There are plenty of red flags indicating that the *Veridical Fakes* video is computer manipulated. For one, Duchamp and Mercury have been dead for decades, long before the creation of social media. There are also visual cues when you look closely at the faces: you can see minute, peculiarly un-humanlike twitches at the edges of Freeman’s face. Trump’s face and neck look as though they are floating on his barely shadowed white dress shirt, and anyone who casually followed American news during his presidency should find that the voice here is a vaguely passable impression. Visual and audio qualities aside, when viewed as a whole, the tone of this video is very parodic, with the celebrities’ mannerisms and public personae being very exaggerated. Zuckerberg brags about stealing your information, Kardashian talks about literally praying to a data-driven god, and then-President Trump boasts about winning the 2016 election in part because of Spectre. Mercury and Duchamp aside, perhaps each clip and quotation could *seem* like something the celebrity would say, but when viewed as a whole, it feels a little too over the top, at least for this discerning viewer.

Falsities versus Lies

It was relatively straightforward to determine that, with regards to a potential truth claim that the video shows the real-life celebrities saying these words to a camera, this is untrue. What is more difficult to discern is if Posters and Howe were even making such a truth claim at all. While we can use other people's perceptions to evaluate the truth of a claim, we know better than to accept all others' accounts unilaterally. People can be mistaken, but they can also lie. When something is determined to be false, this can arouse suspicion of deceit. Bok classifies the search for truthfulness and lying into two broad categories: the moral and the epistemological (6). The moral domain of lying is settled by determining if you *intended* to mislead, and the vaster epistemological domain of truth and falsity is settled by determining truth. She defines deception as a thought process, from which a lie is "an intentionally deceptive message in the form of a statement" (15). Thus, she defines deception as a mental process with the intention to lie, and a lie as the statement. It is possible to say something false if you thought you were telling the truth, but that would not be a lie. For example, if I were to tell you that a mushroom is safe to eat, then you eat it and are poisoned, there are two possibilities: either I knew the mushroom is poisonous and was lying, or I honestly thought that mushroom would be safe to eat and was not lying. In both scenarios, the truth is that the mushroom is poisonous, which is independent of whether or not I was lying.

Of course, this leads to the extra complexity of determining what is a lie, apart from what is simply untrue. Much like the complications of Truth and truth, we are once again restricted by our own perceptions: we do not have access to other people's thoughts and, therefore, intentions. In order to lie, one must intentionally know the truth. How, then, do we know if they knew the truth?

Like Bok, Harry G. Frankfurt is another ethicist who has written about the phenomenon of lying. In his popular 1986 essay, “On Bullshit,” he takes a serious look at what constitutes bullshit with regards to lying. Though seemingly trivial, his distinctions between bullshit and other forms of lying is a useful framework for navigating online misinformation, including potential nefarious uses of ‘deepfakes.’

Frankfurt characterizes a lie as “an act with a sharp focus” (15-16). The liar must be cognizant of a particular truth claim in order to intentionally misrepresent it. He contrasts this with a bullshitter:

On the other hand, a person who undertakes to bullshit his way through has much more freedom. His focus is panoramic rather than particular. He does not limit himself to inserting a certain falsehood at a specific point, and thus he is not constrained by the truths surrounding that point or intersecting it. He is prepared to fake the context as well, so far as need requires. (16)

In other words, a bullshitter is defined by a “lack of connection to a concern with truth” (Frankfurt 10). While a liar is aware that they are not being truthful, that they are making a deliberate effort to conceal the truth, a bullshitter does not regard the truth at all. A liar is making a pointed decision for a precise truth claim, whereas a bullshitter has no intention of either telling the truth or lying: “the motive guiding and controlling [the bullshit] is unconcerned with how the things about which he speaks truly are” (Frankfurt 17). Interestingly, bullshit need not be false; what needs to be false is the bullshitter’s motives (14). The bullshitter does not care if what they are saying is true or not. Their intention is to make them look good in the eyes of others (Frankfurt 16). This style of substance-less self-aggrandizing is best exemplified in politicians. It is easy to draw comparisons between skeptics’ mistrust of truth and this disregard for it:

“convinced that reality has no inherent nature, which he might hope to identify as the truth about things, he devotes himself to being true to his own nature” (Frankfurt 20). One potential downside to militant skepticism is this bullshit mindset; the overemphasis of our own perceptions, which may lead us to conflate it with the truth. As George Costanza once said on *Seinfeld*, “it’s not a lie if you believe it.” This is what bullshitters do: they do not care about what is actually the truth, but they think they can use will power to convince themselves that their own falsehoods are true.

Can one be a bullshitter with regards to how they make and use depictions? I argue yes. While I have spent most of this chapter focusing on the importance of evaluating the truthfulness of what you encounter, there is nonetheless responsibility when making or saying things to avoid misinterpretation. In ethics, there is no universal rubric for determining whether the maker or the receiver is at fault. However, that does not mean this is a frivolous goal, for both sides should take care not to mislead or to misinterpret. Context is one such example.

Context

Regardless of the context in which *Veridical Fakes* is presented (in a gallery or on Instagram, for example), how can we tell if Powers and Howe wanted to deceive us, and to what extent? Did they want us to believe every word these celebrities were saying, and to share this with others as though the events depicted really happened? And if they did not want us to believe this, should they not have made it clearer?

What, then, about the individual segments of the *Veridical Fakes* video? Do they seem more plausible when viewed on their own, without being presented in an exhibit slyly skewering Silicon Valley ethos? Maybe yes. Posters and Howe—working with Canny AI, the same

company that helped develop the Moon Disaster video I discussed in the introduction—made a third ‘deepfake’ clip of Zuckerberg (“Tackling the misinformation epidemic”).



Fig. 6. Bill Posters and Daniel Howe. Still from *I wish I knew...* 2019, Instagram.

In this video, Zuckerberg says, in part, “whoever controls the data, controls the future” (@bill_posters_uk). This video, which Posters posted on his Instagram account in June 2019, garnered wide media attention.²⁰ While this was not the first time Facebook (Instagram’s parent company) came under fire for propagating disinformation, it was noteworthy because the

²⁰ See: O’Sullivan, Donie and Rachel Metz. “A Deepfake Video of Mark Zuckerberg Presents a New Challenge for Facebook.” *CNN*, 11 June 2019, <https://www.cnn.com/2019/06/11/tech/zuckerberg-deepfake/index.html>; O’Neil, Luke. “Doctored Video of Sinister Zuckerberg Puts Facebook to Test.” *The Irish Times*, 12 June 2019, <https://www.irishtimes.com/culture/doctored-video-of-sinister-zuckerberg-puts-facebook-to-test-1.3923352>; Cole, Samantha. “The Mark Zuckerberg Deepfakes Are Forcing Facebook to Fact Check Art.” *Vice*, 14 June 2019, <https://www.vice.com/en/article/pajvq9/mark-zuckerberg-deepfake-facebook-misinformation-moderation-satire>.

disinformation the video was alleged to be spreading concerned the platform’s very own CEO. The company decided to allow the video to remain, despite others calling it an example of misinformation (“Facebook Lets Deepfake”).

While some people called for the removal of the Zuckerberg ‘deepfake’ Instagram video, there is minimal, if any, outcry for *Veridical Fakes* or any other celebrity ‘deepfake’ art by Powers and Howe. Exposure is certainly a factor; the Instagram video was widely shared and discussed in news outlet, though some articles included the added context of Powers’ and Howe’s other ‘deepfake’ videos (“Facebook Lets Deepfake”). With this in mind, I posit that the largest driving factor of why the lone Zuckerberg clip is more upsetting to people than the “synthetic art” compilation video is because of its *context*. These two examples demonstrate a moral complexity in the uses of ‘deepfakes’—complexities which exist regardless of medium or technology. When presented in an art gallery, or when viewed on the artist’s personal website, you are less likely to believe these people are really making such outlandish declarations. But if you stumble upon one of these videos on social media, you might be more easily deceived. That is the power of context. Curiously, context is used to evaluate truthfulness, but it can also be used to easily fabricate falsehoods.

‘Deepfakes’ and “Cheap Fakes”

Despite my insistence that the medium does not define how truthful a depiction is, we nonetheless must consider the amounts of truthful and deceptive uses. This is a direct reflection of the discussions in Chapter 1 on automation’s role in technology. I am not contradicting my argument; rather, this strengthens my assertion that *how* we use technology and media is what determines whether it is used to tell the truth or to lie. I insist that we can use any technology or medium for any purpose, but that does not mean that we do so at equal rates. Contrary to those

who worry about ‘deepfakes’ being used for deception, there is a basic philosophical principle, which has endured over time, that suggests we are unlikely to see widespread malicious ‘deepfakes,’ at least compared to other modes of representation, for one simple reason: they are hard to make. Alarmist accounts of this technology often ignore the amount of time and skill required to make any ‘deepfake,’ let alone one perceptually realistic enough to deceive people.

Take, for example, a series of TikTok ‘deepfakes’ of Tom Cruise that went viral in early 2021, leading to widespread coverage. Certain outlets took the opportunity to talk to the videos’ creator, special-effects artist Chris Ume, about how these depictions were made.²¹ Ume describes to *The Verge* (a technology-focused section of *Vox*) how it took two months just to train the algorithm to study Cruise’s face, which then required days to synthesize each ‘deepfake’ clip (Vincent). Once a ‘deepfake’ clip was generated, Ume still had to make numerous manual adjustments to each video, which required him to make frame-by-frame edits (Vincent). All of this hard work and time was already assisted by the fact Ume was using a target video of a Cruise impersonator, Miles Fisher, who bears a striking resemblance to the actor and who did not need an algorithm to make a convincing voice. This example shows just how complicated it can be to make any ‘deepfake’ depiction, even when you have the time, skills and ideal conditions. Even with an immense number of photographs and videos for the algorithm to analyze—which is one of the reasons why ‘deepfakes’ tend to use celebrities, so often

²¹ Perhaps, having encountered so many examples of ‘deepfakes’ over the past few years, media has begun to take a more moderate approach to covering this technology. See: Vincent, James. “Tom Cruise Deepfake Creator Says Public Shouldn’t Be Worried about ‘One-Click Fakes.’” *The Verge*, 5 Mar. 2021, <https://www.theverge.com/2021/3/5/22314980/tom-cruise-deepfake-tiktok-videos-ai-impersonator-chris-ume-miles-fisher>; Gilbert, David. “Here’s How Worried You Should Be About Those Tom Cruise Deepfakes.” *Vice*, 3 Mar. 2021, <https://www.vice.com/en/article/n7vgm8/heres-how-worried-you-should-be-about-those-tom-cruise-deepfakes>.

photographed—and a doppelgänger to act as the stand-in and voice, Ume still needed to make considerable manual changes to the ‘deepfake’ for maximum perceptual realism.



Fig. 7. Steve Ume and Miles Fisher. Still from @deeptomcruise video. 2021, TikTok.

If Ume really wanted to pretend to be Tom Cruise, or to make any other kind of falsehood, he could have done this using any medium. For example, it is quite simple to study Cruise’s handwriting to make forged letters. Instead of a TikTok account, he could have made a parody Twitter account and easily written tweets as though he were Cruise. He could have sung a song or drew a picture, all of which are media we can use to tell a lie. Fisher was already acting

like he was Cruise, which is effectively no different than what Ume did, which required far more time, skill and money.

In summary, we should not be as concerned about misinformation ‘deepfakes’ because it is far easier to deceive using simpler means. This is the basis of a report by Britt Paris and Joan Donovan, who coin the term “cheap fake” to refer to such simple ways to deceive. Paris and Donovan create an axis that shows how photographic video can be further altered in ways that can be simple (“cheap fakes”) or more sophisticated (‘deepfakes’).

Perhaps the most common way to make a “cheap fake” is to take a public figure’s words out of context. For example, during the 2021 Canadian federal election, then-Minister of Finance Chrystia Freeland posted on Twitter a clip of Conservative leader Erin O’Toole giving an interview that suggests that he supports a full privatization of healthcare (@cafreeland). Journalists were quick to point out that O’Toole’s full interview response is far more sympathetic to public healthcare than the thirty-five-second snippet she shared suggested. Freeland and other opponents of O’Toole did not need to spend the time and money to create a ‘deepfake’ of him calling for private healthcare; it was far simpler to selectively edit his own words to suit their own political agenda.

Another easy method to manipulate video is to slow down or speed up existing footage. A video of American Speaker of the House Nancy Pelosi has been slowed down to make it look as though she is slurring her speech and either drunk or otherwise incapacitated (“Pelosi videos manipulated”). Once again, these are very simple methods for creating convincing misinformation, which is far more plentiful online than any supposed ‘deepfake.’

Summary

Having determined in Chapter 1 that it is possible to lie using any medium or technology because lies are not inherent material qualities, this chapter was dedicated to defining truth and how representations are imbued with this, and other, qualities. I showed how there are two types of truth, of which one is relevant for evaluating truth claims.

The next chapter is a continuation of these questions from the specific lens of AI itself. The existence of AI does not contradict any of the aforementioned argument, but it would be prudent to approach these questions once again, this time from the perspective of those who argue that AI is unprecedented ontologically.

Chapter 3: ‘Decoding’ the Algorithm and New Media

In Chapter 1, I established how all new technologies, ‘deepfakes’ included, confound our past understandings of what technology is. While AI shares the same philosophical concerns that I have discussed at length in the first chapter, including how automation is a prevailing worry with new technology, it is also worth homing in on AI in particular. This is because the sociocultural discourse surrounding ‘deepfakes’ is a direct reflection of how we characterize AI, a broad group of technologies that includes the algorithms used to make ‘deepfakes.’ Additionally, I feel compelled to address AI on its own, not merely under the wider umbrella of technology, because there is prevailing discourse that treats AI as an exceptional mode of technology, one whose ramifications are unlike any other we have ever seen before. I want to interrogate these claims in this chapter.

We need not concern ourselves with the minutia of the medium itself, since, as I argue in Chapter 1, different media and technologies are more similar to one another than dissimilar. This chapter is not for addressing whether or not AI-based technologies, including the ‘deepfake,’ are materially deceptive, as I dedicate the previous two chapters to this inquiry more broadly. Furthermore, I have already argued in favour of an actions-based model for characterizing truth in Chapter 2; this means that AI can be used for any kind of discursive use, including telling the truth and lying. The goal of this chapter is to expand upon the technological concepts of Chapter 1, as these concepts specifically apply to AI. I describe the common arguments that clash with my account of truth, lies and ‘deepfakes’: that its use is an abrupt departure from our previous relationship to other technologies, that the effects of AI use is unprecedented, and that such digital representations hold more information and are easier for deception.

I begin this chapter by offering a brief history of computer-generated faces. The reason for this is to demonstrate that, as with most types of technologies, their development is neither linear nor simple. This will help disprove some of the sociocultural concerns surrounding ‘deepfakes’. We have been using computers to alter video faces since around the mid-twentieth century, and so the emergence of ‘deepfakes’ in the public conscience was not without precedence.

I then offer cursory definitions of AI and algorithms to situate us. I refer to the relatively new subfield of algorithm studies for valuable insight into our contemporary attitudes toward AI and algorithms. In what ways are these attitudes toward ‘deepfakes’ a manifestation of our fear of AI? This is a logical extension of our fears of automation, a common concern surrounding all technologies, as discussed in Chapter 1. I maintain that AI and ‘deepfakes’ are no different than any other technology, as any new technology leads to a critical re-evaluation of our understandings of technology as a whole.²² Nonetheless, I devote this chapter to discussing AI and algorithms as particular instances of technology, given their use in ‘deepfakes.’

Finally, I discuss the concepts of digital and analog. I show how our binary ways of classifying representations as either wholly digital or analogical is fundamentally flawed, as all images contain both. Closely related to this digital-analog dichotomy is the structuralist movement, which I briefly discuss, as it is commonly employed in modern accounts of AI. I use the distinctions of digital and analog, as well as material determinism, to argue against a distinction in amounts of information in images, as well as its connection to truthfulness.

²² Recall in Chapter 1 that Guillory illustrates this using the advent in the nineteenth-century of the telegraph and phonograph, which were then considered “new media.”

Brief History of Computer-Generated Faces

I want to begin by considering the history of visual effects, particularly for creating video faces using computers. The history of computer-generated faces spans decades. In lieu of an exhaustive genealogy, I will provide a few key examples to demonstrate the trajectory of inventions leading to the creation of the ‘deepfake,’ paying special attention to methods of generating faces using computers. Much like in the first chapter, I am once again emphasizing the action of creating a photographic representation. A critical look at the history of making faces using cameras and computers demonstrates that the ‘deepfake’ is not an outlier in our continuous re-evaluation of what is possible with photography. As you will see, these techniques all rely on photography’s ability to mark surfaces using light. This means that, per Maynard’s account, even computer-generated faces are simultaneously detective and depictive.

Motion Capture

In order to understand the rise of ‘deepfakes’ and other methods of transforming photographically captured faces into synthetic ones, I first need to discuss motion-capture technology, which is limited to the body and not facial features. This technology, commonly referred to as “mo-cap,” first allowed visual-effects artists to transform photographically captured movement into computer-generated images (CGI). It was after having used this technology that artists were eventually able to capture the more minute facial expressions through similar technologies.

An early example of this motion-capture technology for the body is Scanimate. First developed by Lee Harrison III in 1967, this is a technology in which the “human operator interactively [moves] layered bits of imagery around on a TV screen and then [plays] back the motions to videotape” (Sturman 38). Put simply, Scanimate involves making an ephemeral

electronic image that is captured on tape. Best remembered for the distinct animations of TV station logos in the 1970s and 1980s, Scanimate had multiple uses and effects, depending on how it was operated. Initially, Harrison used dials and knobs to interact with the imagery, before successfully equipping a dancer into a special suit equipped with sensors (Sturman 38).

While mainly restricted to academia and tech tradeshows in the latter half of the twentieth century,²³ the 2000s saw a sharp increase in the use and development of motion-capture technology in popular video games and cinema. There are currently a few different types of motion-capture suits. If using optical or infrared sensors, the actor is rigged with special light-reflecting or light-emitting markers, strategically placed on actor's body (Prince, *Digital Visual Effects* 119-120). The actor is positioned around multiple cameras that capture the vertices of lights reflected or emitted from the sensors, allowing for three-dimensional modelling when combining the data from all cameras (Prince, *Digital Visual Effects* 120).

Memorably, the character Gollum in Peter Jackson's *The Lord of the Rings* trilogy (2001, 2002 and 2003) demonstrates the latest technology for character animation in the early aughts. A motion-capture suit was used to translate actor Andy Serkis's movements into an animatable creature, though, as had been the case with this particular technology, his face was not rigged to do the same (Allison, "The Modern Entertainment Marketplace" 177; Prince, *Digital Visual Effects* 120).

Performance Capture and Facial Mapping

It was not until Serkis and Jackson worked together again on *King Kong* (2005) that motion-capture sensors would be placed on his face (Prince, *Digital Visual Effects* 120). Creating markers

²³ Notably the SIGGRAPH annual conference, where a rudimentary example of facial motion capturing was debuted with Mike the talking head in the late 1980s (Sturman 38-39).

allows for facial mapping, which is foundational for computer manipulating faces. This combination of sensors on the face and body is a specific type of motion capture called performance capture (Prince, *Digital Visual Effects* 120). There was an added challenge in transforming a human actor into another animal, however. Despite Serkis's best efforts to act apelike, there are some jaw and upper-lip movements gorillas make that humans are physiologically unable to mimic. To counteract this, an algorithm was 'taught' how to 'translate' different movements, adjusting Serkis's face onscreen when he would make particular facial expressions (Allison, "More than a Man" 329). This technology still required frequent manual corrections, as it is "a messy process, rife with illegible data, inconsistencies, and errors" (Allison, "More than a Man" 329). This is an example of using an algorithm to manipulate a face onscreen, which, like 'deepfakes,' requires input from photographically captured imagery.

King Kong used performance capture, in which Serkis's body and face were tracked, whereas *Avatar* (James Cameron, 2009) only used motion-capture bodysuits. Instead of sensors on the actors' faces, they wore helmets mounted with mini-cameras that simultaneously filmed their faces from different angles (Prince, *Digital Visual Effects* 120). In post-production, "the animators numerically encoded every facial muscle and created a character rig with this data that could produce minute responses such as an eyelid twitch or an upper-brow raise," transforming the low-resolution images from the helmets into more perceptually realistic faces (Prince, *Digital Visual Effects* 135).

A similar technique was later employed in *The Irishman* (Martin Scorsese, 2019). As a film whose timeline alternates between decades, the filmmakers opted for computer effects to de-age the lead characters, instead of using makeup or casting for the characters' different ages. Like *Avatar*, the visual effects team relied on three cameras to create a three-dimensional model

(Helman). Where *The Irishman* differs from *Avatar* is that there were no cameras attached to the actors. Instead, the two additional cameras were placed on either side of what they dubbed “the director’s camera” (Helman). Instead of relying on placed facial markers to construct a three-dimensional model on a computer, the software used for *The Irishman* “gave [filmmakers] all the pixels that the camera sees in somebody’s face. Now, instead of 200 markers, you have thousands of markers, constantly moving” (Helman). This technique—which allowed Robert DeNiro to look twenty years younger without applying heavy makeup, wearing obtrusive helmet cameras, or placing markers all over his face—is similar to how ‘deepfakes’ are made.

The Irishman did not use actual ‘deepfake’ technology, though, because the filmmakers were not trying to synthesize another person’s face on top of the actors, nor were they trying to alter the words the actor is saying. Regardless, the technologies are very similar in practice, with one other big difference. For a ‘deepfake,’ we are unlikely to have access to three videos of the original face taken at different camera angles. To compensate for this lack of three dimensionality, the algorithm is provided with many photos and videos of the target face, ideally taken from as many angles as possible. Thus, while we do not have a three-dimensional model of the original face, as they did for *The Irishman*, the algorithm is able to construct one for the target face, achieving similar results.

In both cases, the technology relies on the ability to map facial features. The algorithm’s analysis of human faces—identifying contours and movements when talking or making certain expressions, even recognizing differences between faces, as is the case with facial-recognition technology—is a growing worry for many. Though this thesis is predicated on the assumption that the epistemic threats we face are not unique, but rather something we have had to contend with since the very beginning of communication, we are still very fearful of AI. As explained above,

the creation of computer-generated faces spans decades. And yet, there is a palpably stronger concern with AI in particular, due to the speed and perceived lack of human agency required.

Structuralism and Digital Media

One of the five principles that Lev Manovich uses to describe new media is that they are all represented numerically. “All new media objects,” he says, “whether created from scratch on computers or converted from analog media sources, are composed of digital code; they are numerical representations” (27). This, he argues, has two key consequences: one, a new media object can be reduced to a mathematical formula; and two, we can manipulate new media using algorithms. These algorithmic manipulations include “automatically [removing] ‘noise’ from a photograph, [improving] its contrast, [locating] the edges of the shapes, or [changing] its proportions” (Manovich 27). Another such principle from Manovich is that new media are highly variable (36). He contrasts old media, which is “stored in some material, its order determined once and for all,” with the relative ease of refiguring new media (Manovich 36). Manovich is thus arguing that such new media are inherently more capable of alteration. One consequence of this increased ability for refiguration, according to this logic, is an increased ability to deceive.

This is a tenet of Lev Manovich’s effectively structuralist argument, a movement that suggests objects and concepts are best understood if we divide them into discrete units. Manovich denies being a structuralist, writing that despite the implications of his book title *The Language of New Media*, “I do not want to suggest that we need to return to the structuralist phase of semiotics in understanding new media” (12). Despite this protestation, there are nonetheless structuralist implications to his argument, particularly that there is now a singular code that underlies all representation. I do not have sufficient time or space here to address the complexities of semiotics in cinema, though it relates to structuralism. That said, I echo

Maynard's sentiment that, for our purposes, it is important to consider uses of visual displays, in part because "a word like 'sign' is certain to obscure this point, suggesting a mythical common function for all such markings" (60). Too often does semiotics ask us to restrict our analysis to a singular meaning, a kind of logic that leads to the kind of essentialist accounts of media that I argue against in Chapter 1.

The Analog and the Digital

When discussing the development of new technologies, we often refer to analog and digital, but seldom do we take the time to consider the distinction when discussing film. Manovich points to three different ways of characterizing a type of film as digital: "analog-to-digital conversion (digitization), a common representational code, and numerical representation" (52). It is the ambiguity of distinguishing these terms that has led scholars like Manovich away from the broad categorization of "analog film" and "digital film," and instead to "old media" and "new media" (52). While I think it is useful to move away from "analog" and "digital" as blanket terms—on which I will elaborate shortly—"new media" is still a frustratingly vague term. For one, there is no clear consensus on how new it is. The histories of the camera and the computer have run parallel to one another since the mid-nineteenth century (Manovich 22). There was never an "inherent discontinuity cleaving the digital from the analogical arts" with regards to how films were made (Rodowick vii). As demonstrated with the abridged history of computer-generated faces, digital use in film predates the mass adoption of digital cameras, for instance.

These terms, "analog" and "digital," have been used in so many different contexts that they have begun to lose meaning. Nelson Goodman noted long ago that "the difference between

analog and digital machines or systems is easier to illustrate than to define” (159).²⁴ I want to centre my argument in the broad philosophical concepts of the digital and the analog. Doing so will prevent the overly simplified characterizations of technologies that have become pervasive. Before Goodman, philosopher David Lewis distinguished “analog” and “digital” based on whether or not the representation can be differentiated (321). While digital representation can be divided into discrete units, analog representation is too dense for simple division. However, Lewis concedes that analog representation can also be numerically divided. He uses the example of measuring voltage, where the ‘analog’ physical magnitude of voltage is translated into ‘digital’ numerical formatting. The true difference between analog and digital, then, is “by physical magnitudes that are either primitive or almost primitive,” according to his mathematical definitions (325).

Though this thesis is not directly engaged with such complex mathematics, Lewis’s definition shows the flexibility in characterizing digital and analog: if we look closely enough, we can make any analog representation digital, thus undermining the importance of these distinctions. It should not be an all-or-nothing approach; we can see nuance in classifying representations as analog or digital, though, ultimately, there is a point at which the digitality of a representation is not sufficient enough to be considered digital. I do not intend to be flippant

²⁴ Goodman offers several examples to illustrate this, saying that a “pressure gauge . . . is a pure and elementary example of what is called an analog computer. The dime-counter displaying numerals is a simple example of what is called a digital computer; and an ordinary watch, read in the most usual way, combines analog and digital computers” (159-160). Despite such obvious examples, though, he says about the analog-digital distinction that “some current notions about it are mistaken. Plainly, a digital system has nothing special to do with digits, or an analog system with analogy. The characters of a digital system may have objects or events of any kind as their inscriptions; and the compliants under an analog system may be as remote and different as we please from the characters” (160). Despite being “misleading,” Goodman says, the terms analog and digital are “unlikely to be discarded.” He suggests that “perhaps the best course is to try to dissociate them from analogy and digits and a good deal of loose talk, and distinguish them in terms of density and differentiation—though these are not opposites” (160). Writing in 1976, Goodman was trying to clarify the distinction as part of his theory of symbols. Arguably, the talk about the distinction between analog and digital has only gotten looser. See Goodman, Nelson. *Languages of Art: An Approach to a Theory of Symbols*. Hackett, 1976.

toward the distinction between analog and digital in general; rather, I want to be clear that whether or not a representation is considered one or the other has little bearing on its ontology, much less its veridical quality or its deceptiveness, since there is no clear demarcation between the analog and the digital.²⁵

Human Agency and Algorithms

Perhaps the biggest misconception with AI is that there is little to no human involvement required. As I hope my philosophical account of technology (drawing heavily from that of Maynard) has convinced you, the value of technology necessarily depends on how humans use it. This is no different with a pencil as it is with sophisticated algorithms used to generate video faces. Unfortunately, much like how the term “deepfake” falsely equates the technology with duplicity, the term “artificial intelligence” (and the related terms of “machine learning” and “deep learning”) reflects the misconception of these technologies as having greater independence from human control, even if we do not consider them to be sentient.²⁶ Thus, we need a proper understanding of the main AI technology used in ‘deepfakes’: algorithms.

Steven S. Skiena defines an algorithm as “a procedure that takes any of the possible input instances and transforms it to the desired output” (Skiena 3). He notes that algorithms need to solve *instances of a problem*, rather than the problems themselves (3). This linguistic choice

²⁵ Often, though, specific values have been attached to the two terms in the distinction. Anthony Wilden, for example, argued in 1980 that: “In human communication, translation from the analog to the digital often involves a gain in information (organization) but a loss in meaning. Translation from the digital to the analog . . . usually involves a loss of information and a gain in meaning” (168). However, Wilden seems to be reducing the complex notion of “meaning” to a purely formal process. These confusions persist in contemporary accounts of digital and analog representation. See Wilden, Anthony. *System and Structure: Essays in Communication and Exchange*, 2nd edition. Tavistock, 1980.

²⁶ For a critique of the idea that machines could be sentient, see Fetzer, James H. *Computers and Cognition: Why Minds are Not Machines*. Kluwer, 2001. At least in its “strong” form, the claim that computers, or AI, “actually possess mentality when they are executing programs,” Fetzer insists that “we have found nothing that would enhance the credibility of the thesis of strong AI” (159, 160).

emphasizes the algorithm's lack of human agency: the algorithm only treats a series of inputs as a problem to solve if it has been programmed to see it as such. James Fetzer makes a similar point, distinguishing between computer program and algorithm, arguing that a "program . . . is a particular implementation of an algorithm in a form that is suitable for execution by a machine", effectively contrasting it with any more complex set of instructions that would require a human to interpret them (204). Human agents are required to devise (and identify) algorithms that are machine-executable, and in this respect no computer system could operate independently of human control. The human interacting with the algorithm needs to tell it what output it wants, whether it is to sort names, to solve math equations, or to construct synthetic faces, as is the case for 'deepfakes.' It is crucial to keep in mind this fundamental relationship between algorithms and humans; doing so will prevent hyperbolic reactions to what can be an otherwise relatively mundane type of technology. Algorithms should not be seen as harbingers of the technological apocalypse, nor as wholly benevolent tools that are estranged from human error and bias.

Lorena Jaume-Palasi argues that AI, rather than being sentient or even material, is a new kind of infrastructure:

It is not a product; it is *immaterial* infrastructure. AI is technology that standardizes and automates processes. Everything that is a process implies a certain system and a set of standards, which can then be formalized in mathematical language and become partially or fully automatable. (479, emphasis in original)

Returning to discussions of medium and technology in the first chapter, Jaume-Palasi and I agree that AI does not inherently imbue particular qualities in its representations. While AI, like any technology, amplifies and suppresses our powers to do things, these outcomes are either positive or negative depending entirely on human value judgments.

Jaume-Palasi may *appear* to be equating this technology with negative qualities—which I argue against—because she later says, “algorithmic systems and artificial intelligence as collectivistic technologies amplify a weakness of democracies” (495). However, what she is actually arguing is that these technologies lay bare what were already structural flaws to this political system, not that they bear any responsibility for this weakness.

Similarly, Paris and Donovan argue that ‘deepfakes,’ when treated as unique forms of deception that require further intervention, can lead to power imbalances: “put another way, new media technologies do not inherently change how evidence works in society. What they do is provide new opportunities for the negotiation of expertise, and therefore power” (22). We do not need to reinvent how to evaluate truth claims, and to suggest otherwise—for example through mandated labelling or using ‘deepfake’-detecting technology²⁷—could have disastrous effects for society. I want us to instead treat labels or detection technology as we would any other criteria that we use in assessing the truthfulness of a claim. This means we should not heedlessly accept all labels or ‘deepfake’ detector assessments as definitive proof of truthfulness.

Summary

This chapter is a summation of three broad categories of inquiry regarding the characterization of ‘deepfakes,’ all with a particular emphasis on AI. In the first section, I focus on the technologies that have led to the development of ‘deepfakes,’ which demonstrate that we have been using computers to alter photographically captured imagery for decades, while still

²⁷ One such example is from December 2019, when Meta AI (then known as Facebook AI) launched the Deepfake Detection Challenge, which encouraged computer engineers to build technology that could best detect digitally manipulated media. The top detector had a 65.18% accuracy for the black box (i.e., closed source) dataset (“Deepfake Detection Challenge Results”). See “Deepfake Detection Challenge Results.” *Meta AI* [Formerly *Facebook AI*], 12 June 2020, <https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai/>.

making ties to its detective properties. Finally, I return to the question of automation that I first posed in Chapter 1, this time focusing on the specifics of ‘deepfakes’ and their use AI. I describe how algorithms necessarily require human agency in order to function, and that their use’s implications are only as dangerous as the people who use it.

Conclusion

This thesis has been an effort to convince readers that ‘deepfakes’ can be used truthfully or deceptively. Truth and falseness are not inherent to any technology or medium, for these are human-ascribed qualities that are highly contextual in nature. When referring to truth, we are specifically referring to a particular truth claim, not any all-encompassing notion of reality. These truth claims can be evaluated using our perceptions, as well as trusting the perceptions of others. This is as good an understanding of truth as we will ever have, given that our experience of the world is invariably subjective. The fact that ‘deepfakes’ are created using AI does not make them any different than other technologies in the many ways, both positive and negative, to which they can be used.

Not all Doom and Gloom

There are certain uses for ‘deepfakes’ that could qualify as positive, or at least as non-deceptive. As illustrated through the brief history of CGI faces in Chapter 3, filmmakers can use this kind of technology to enhance characters’ features. I hope I have made a compelling argument that we should not worry about ‘deepfakes’ in particular deceiving us, the reasons for this being twofold: firstly, no film should be equated with truthfulness, since that is a quality humans ascribe and not an inherent material quality; and secondly, for films in which its fictionality is obvious, we only focus on the detective function of film insofar as it allows for depiction. Like any other tool, ‘deepfakes’ can be used to spread lies, but, more importantly, there is a myriad of good for which this technology can be used.

‘Deepfakes’ can be used for harmless entertainment, as is the case with Ctrl Shift Face, a YouTube channel that posts comedic videos of pop culture. Ctrl Shift Face’s most popular videos are of ‘deepfakes’ to accentuate comedian Bill Hader’s impressions of Arnold

Schwarzenegger, Al Pacino and Tom Cruise. These videos, which use the original audio from Hader, rapidly alternate between Hader's face and that of the actor he impersonates, demonstrating the relative fluidity of both the technology and Hader's acting. The channel also has a series of videos where a 'deepfake' Jim Carrey plays the lead in *The Shining* (Stanley Kubrick, 1980). This juxtaposition of a character actor in a memorable role is a common use of amateur 'deepfakes,' creating absurdity and allowing us to imagine test reels for unusual casting choices.

There can also be positive uses for people being filmed, when providing testimony, for example, but who want to protect their identity, as is the case in *Welcome to Chechnya* (David France, 2020). This documentary focuses on activists who rescue gay men from Chechnya, a Russian republic whose government has been systematically purging men on the basis of sexual orientation. Given the sensitive nature of the men fleeing persecution, and the very real threat of harm for speaking out, the filmmakers have used 'deepfake' technology to alter their faces. Audiences are still able to see nuances in the men's facial expressions as they speak about their experiences, without having to shoot only from behind or using darkly-lit shots of their fronts. It is easy to tell which faces are 'deepfakes,' since there is a slight white glow around the edges. The 'deepfakes' are perceptually realistic enough to resemble faces, while still being noticeably constructed.

This use of 'deepfakes' to mask vulnerable populations (without the expense of showing faces onscreen) can also be used in fictional contexts. For instance, children can be spared of potentially traumatizing filming conditions by using adult stand-ins whose faces are de-aged using 'deepfake' technology. Animals are another vulnerable group that has already benefited from computer technology, which enhances filmmakers' depictive functions with less reliance on

the detective. Instead of stressing, traumatizing or injuring animals on film sets, big-budget studios have used CGI.²⁸

Where to Go from Here

In asking these large philosophical questions, my more specific goal with this thesis was not just to rehabilitate the reputation of ‘deepfakes.’ In doing so, I want to draw our attention to serious lapses in our popular understanding of technology and knowledge. I do not claim to have definitively accounted for the curious nature of these massive concepts, which will continue to elude us for as long as we care to ponder these philosophical questions. What I do want you to have gained here is a sound framework for understanding how we interact with all depictions, including ‘deepfakes,’ in a consistent manner. I wish to emphasize that what I have provided is simply *a* framework, for it is always possible that new, radical developments in technology could confound my account. ‘Deepfakes’ are a relatively new phenomenon, a mere blip on our timeline of video depictions. However, my research reflects a basic understanding of the philosophical discourse that follows any new technology, which has proven to be very cyclical in nature. Nonetheless, as I indicate in my discussion of the two kinds of truth, I can only base my framework on the past and present; I do not have access to the all-knowing capital ‘t’ Truth, which could, theoretically, also encompass the future.

For a long time, “video was the last frontier—the one medium consumers could watch and not automatically assume it could be faked” (Hancock and Bailenson 150). This has led to an overconfidence that film is inherently truthful, in part because we have mistakenly given the

²⁸ For literature on animal welfare in filmmaking, see Parker-Starbuck, Jennifer. “Animal Ontologies and Media Representations: Robotics, Puppets, and the Real of ‘War Horse.’” *Theatre Journal*, vol. 65, no. 3, The Johns Hopkins University Press, 2013, pp. 373–393.

medium these properties. While I have stated my concerns with the post-photography movement's skepticism,²⁹ I agree with Mitchell that we should see the continued proliferation of digital representation as "a welcome opportunity to expose the aporias in photography's construction of the visual world, to deconstruct the very ideas of photographic objectivity and closure, and to resist what has become an increasingly sclerotic pictorial tradition" (8). I do not wish for us to treat into a defeatist attitude of mistrust by default, but rather of neutrality until provided compelling evidence for or against particular truth claims.

'Deepfakes' are not the only examples of widespread mistrust and misinformation. Of course, there are vast sociopolitical implications that lead to mass cynicism and the rise of conspiracy theories. But on a fundamental level, our uneasy relationship with depictions has contributed to this malaise. Suppressing 'deepfakes' will not meaningfully deter misinformation, since we are able to lie using any medium, to the same effect. Truth and lying are too complicated to treat 'deepfakes,' or any other medium or technology, as the scapegoat. The proper course of action against misinformation (aside from prevention through moral appeals) is to use the same judgments we have been using for centuries.

²⁹ See Chapter 1.

Bibliography

“About.” *First Draft*, <https://firstdraftnews.org:443/about/>. Accessed 27 July 2021.

Allison, Tanine. “The Modern Entertainment Marketplace, 2000-Present: Special/Visual Effects.” *Editing and Special/Visual Effects*. Edited by Charli Keil and Kristen Whissel. Rutgers University Press, 2016, pp.172-186.

_____. “More than a Man in a Monkey Suit: Andy Serkis, Motion Capture, and Digital Realism.” *Quarterly Review of Film and Video*, vol. 28, no. 4, Routledge, July 2011, pp. 325–41.

@bill_posters_uk. Video of *I wish I could...* (2019). *Instagram*, 13 June 2019, <https://www.instagram.com/p/BypkGIvFfGZ/>.

@cafreeland. Video of Erin O’Toole. *Twitter*, 22 Aug. 2021, <https://twitter.com/cafreeland/status/1429429537012658178>.

Bazin, André. “The Ontology of the Photographic Image.” *What is Cinema?* Translated by Hugh Gray. University of California Press, 1971 [1958], pp. 9-16.

Blackburn, Simon. *Truth: A Guide*. Oxford University Press, 2005.

Bok, Sissela. *Lying: Moral Choice in Public and Private Life*. Random House, 1978.

BuzzFeedVideo. “You Won’t Believe What Obama Says In This Video! 😊.” *YouTube*, https://www.youtube.com/watch?v=cQ54GDm1eL0&ab_channel=BuzzFeedVideo. Accessed 25 July 2021.

“Camera Obscura, n.” *OED Online*, Oxford University Press. *Oxford English Dictionary*, <http://www.oed.com/view/Entry/272777>. Accessed 14 Dec. 2021.

Carroll, Noël. “Film/Mind Analogies: The Case of Hugo Munsterberg.” *The Journal of Aesthetics and Art Criticism*, vol. 46, no. 4, 1988, pp. 489-499.

- _____. “Forget the Medium!” *Engaging the Moving Image*. Yale University Press, 2003, pp. 1-9.
- Cavell, Stanley. *The World Viewed: Reflections on the Ontology of Film*. Harvard University Press, 1979.
- Cole, Samantha. “AI-Assisted Fake Porn Is Here and We’re All Fucked.” *Motherboard*, 12 Nov. 2017, <https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn>.
- _____. “The Mark Zuckerberg Deepfakes Are Forcing Facebook to Fact Check Art.” *Vice*, 14 June 2019, <https://www.vice.com/en/article/pajvq9/mark-zuckerberg-deepfake-facebook-misinformation-moderation-satire>.
- Cross, Mai’a K. Davis. “The Social Construction of the Space Race: Then and Now.” *International Affairs*, vol. 95, no. 6, Nov. 2019, pp. 1403–21.
- Ctrl Shift Face’s Channel. *YouTube*, https://www.youtube.com/channel/UCKpH0CK1tc73e4wh0_pgL3g. Accessed 8 Oct. 2021.
- “Deepfake Detection Challenge Results.” *Meta AI [Formerly Facebook AI]*, 12 June 2020, <https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai/>.
- Dunne, Matthew W. *A Cold War State of Mind: Brainwashing and Postwar American Society*. University of Massachusetts Press, 2013.
- “Facebook Lets Deepfake Zuckerberg Video Stay on Instagram.” *BBC News*, 12 June 2019, <https://www.bbc.com/news/technology-48607673>.
- Fallis, Don. “The Epistemic Threat of Deepfakes.” *Philosophy & Technology*, vol. 34, no. 4, 2020, pp. 623-643.

“Fenton Crimean War photograph collection.” *National Archives*, n.d.,

<https://www.loc.gov/pictures/item/2001696100/>.

Fetzer, James H. *Computers and Cognition: Why Minds are Not Machines*. Kluwer, 2001.

Finn, Jonathan M. “Images, Objectivity, and Visual Evidence in Sport.” *Beyond the Finish Line: Images, Evidence, and the History of the Photo-Finish*. McGill-Queen’s University Press, 2020, pp. 3-25.

Frankfurt, Harry. “On Bullshit.” *Raritan Quarterly Review*, vol. 6, no. 2, 1986, pp. 81–100.

Friedel, Robert N. *A Culture of Improvement: Technology and the Western Millennium*. MIT Press, 2007.

Gilbert, David. “Here’s How Worried You Should Be About Those Tom Cruise Deepfakes.”

Vice, 3 March 2021, <https://www.vice.com/en/article/n7vgm8/heres-how-worried-you-should-be-about-those-tom-cruise-deepfakes>.

Goodman, Nelson. *Languages of Art: An Approach to a Theory of Symbols*. Hackett, 1976.

Gow, Glenn. “The Scary Truth Behind The FBI Warning: Deepfake Fraud Is Here And It’s Serious—We Are Not Prepared.” *Forbes*, 2 May 2021, <https://www.forbes.com/sites/glenngow/2021/05/02/the-scary-truth-behind-the-fbi-warning-deepfake-fraud-is-here-and-its-serious-we-are-not-prepared/>.

Guillory, John. “Genesis of the Media Concept.” *Critical Inquiry*, vol. 36, no. 2, The University of Chicago Press, 2010, pp. 321–62.

Hancock, Jeffrey T., and Jeremy N. Bailenson. “The Social Impact of Deepfakes.”

Cyberpsychology, Behavior, and Social Networking, vol. 24, no. 3, Mary Ann Liebert, Inc., publishers, 2021, pp. 149–52.

- Harwell, Drew. “Remember the ‘deepfake cheerleader mom’? Prosecutors now admit they can’t prove fake-video claims.” *Washington Post*, 14 May 2021, <https://www.washingtonpost.com/technology/2021/05/14/deepfake-cheer-mom-claims-dropped/>.
- Heidegger, Martin. “The Question Concerning Technology.” *Martin Heidegger: Basic Writings*. Edited by David Farrell Krell. Harper, 1993 [1949].
- Helman, Pablo. “How The Irishman’s FX Team Used Bleeding-Edge Tech to De-Age De Niro.” Interview by Rick Marshall. *Digital Trends*, 5 Feb. 2020, <https://www.digitaltrends.com/movies/the-irishman-netflix-visual-effects-interview-digital-de-aging-robert-de-niro/>.
- Hern, Alex. “‘Deepfake’ Face-Swap Porn Videos Banned by Pornhub and Twitter.” *The Guardian*, 7 Feb. 2018, <http://www.theguardian.com/technology/2018/feb/07/twitter-pornhub-ban-deepfake-ai-face-swap-porn-videos-celebrities-gfycat-reddit>.
- “In Event of Moon Disaster.” *National Archives*, 15 Aug. 2016, <https://www.archives.gov/presidential-libraries/events/centennials/nixon/exhibit/nixon-online-exhibit-disaster.html>.
- Jaume-Palasi, Lorena. “Why We Are Failing to Understand the Societal Impact of Artificial Intelligence.” *Social Research: An International Quarterly*, vol. 86, no. 2, 2019, pp. 477-498.
- Kelion, Leo. “Reddit bans deepfake porn videos.” *BBC News*, 7 Feb. 2018, <https://www.bbc.com/news/technology-42984127>.
- Kracauer, Siegfried. “Photography.” *Theory of Film: The Redemption of Physical Reality*. Princeton University Press, 1960, pp. 3-40.

- Krell, David Farrell. "Heidegger." *Martin Heidegger: Basic Writings*. Edited by David Farrell Krell. Harper, 1993.
- Landesman, Charles, and Roblin Meeks, Editors. "Plato, from *Apology*." *Philosophical Skepticism*. Blackwell Publishing, 2003, pp. 9-22.
- Lees, Dominic, et al. "The Digital Resurrection of Margaret Thatcher: Creative, Technological and Legal Dilemmas in the Use of Deepfakes in Screen Drama." *Convergence: The International Journal of Research into New Media Technologies*, Aug. 2021, pp. 954-973.
- Lewis, David. "Analog and Digital." *Noûs*, vol. 5, no. 3, Wiley, 1971, pp. 321–27.
- Lynch, Michael P. *True to Life: Why Truth Matters*. MIT Press, 2004.
- Manovich, Lev. *The Language of New Media*. MIT Press, 2001.
- Marvin, Carolyn. *When Old Technologies Were New: Thinking About Electrical Communication in the Late Nineteenth Century*. Oxford University Press, 1988.
- Maynard, Patrick. *The Engine of Visualization: Thinking through Photography*. Cornell University Press, 1997.
- Mitchell, William J. *The Reconfigured Eye: Visual Truth in the Post-Photographic Era*. MIT Press, 1992.
- Morgan, Daniel. "Rethinking Bazin: Ontology and Realist Aesthetics." *Critical Inquiry*, vol. 32, no. 3, 2006, pp. 443-481.
- Morris, Errol. "Which Came First, the Chicken or the Egg?" *Believing is Seeing*. The Penguin Press, 2011, pp. 3-71.
- Murdoch, Iris. "Art is the Imitation of Nature." *Existentialists and Mystics: Writings on Philosophy and Literature*. Penguin Books, 1997 [1978], pp. 243-257.

“National Aeronautics and Space Act of 1958 (Unamended).” *NASA*,
<https://history.nasa.gov/spaceact.html>.

“New Media, n.” *Oxford English Dictionary*. <https://www-oed-com.proxy.library.carleton.ca/view/Entry/255819?redirectedFrom=new+media#eid>.
 Accessed 24 July 2021.

O’Neil, Luke. “Doctored Video of Sinister Zuckerberg Puts Facebook to Test.” *The Irish Times*,
 12 June 2019, <https://www.irishtimes.com/culture/doctored-video-of-sinister-zuckerberg-puts-facebook-to-test-1.3923352>.

O’Sullivan, Donie. “Deepfake Videos: Inside the Pentagon’s Race against Disinformation.”
CNN, <https://www.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/>. Accessed 12 Feb. 2021.

O’Sullivan, Donie, and Rachel Metz. “A Deepfake Video of Mark Zuckerberg Presents a New
 Challenge for Facebook.” *CNN*, 11 June 2019,
<https://www.cnn.com/2019/06/11/tech/zuckerberg-deepfake/index.html>.

“Pelosi videos manipulated to make her appear drunk are being shared on social media.”
 YouTube, uploaded by Washington Post, 23 May 2019,
<https://www.youtube.com/watch?v=sDOo5nDJwgA&t=50s>.

Peters, John Durham. *Speaking into the Air: A History of the Idea of Communication*. University
 of Chicago Press, 1999.

Panetta, Francesca, and Halsey Burgund. “In Event of Moon Disaster - Behind the Scenes.” *In
 Event of Moon Disaster*, <https://moondisaster.org/explainer>. Accessed 17 July 2021.

Paris, Britt, and Joan Donovan. "Deepfakes and Cheap Fakes." *Data & Society*, Data & Society Research Institute, 18 Sept. 2019, <https://datasociety.net/library/deepfakes-and-cheap-fakes/>.

Parker-Starbuck, Jennifer. "Animal Ontologies and Media Representations: Robotics, Puppets, and the Real of 'War Horse.'" *Theatre Journal*, vol. 65, no. 3, The Johns Hopkins University Press, 2013, pp. 373–393.

Plato. *The Republic*. *Plato: Complete Works*. Edited by John M. Cooper. Hackett, 1997.

Posters, Bill. "'Spectre' (2019)." *Bill Posters*, <http://billposters.ch/projects/spectre/>. Accessed 4 Oct. 2021.

_____. *Veridical Fakes* (2020). *Bill Posters*, <http://billposters.ch/projects/veridical-fakes-2020/>. Accessed 29 Aug. 2021.

_____. "The Art of Interrogation: An Interview with Bill Posters." *Juxtapoz Magazine*, 19 June 2020, <https://www.juxtapoz.com/news/street-art/the-art-of-interrogation-an-interview-with-bill-posters/>.

Prince, Stephen. *Digital Visual Effects in Cinema: The Seduction of Reality*. Rutgers University Press, 2012.

Quine, W.V. *Word and Object*, new edition. MIT Press, 2013 [1960].

Robertson, Adi. "Reddit Bans 'Deepfakes' AI Porn Communities." *The Verge*, 7 Feb. 2018, <https://www.theverge.com/2018/2/7/16982046/reddit-deepfakes-ai-celebrity-face-swap-porn-community-ban>.

Rodowick, D. N. *The Virtual Life of Film*. Harvard University Press, 2007.

Rothman, William, and Marian Keane. *Reading Cavell's The World Viewed: A Philosophical Perspective on Film*. Wayne State University Press, 2000.

Scruton, Roger. "Photography and Representation." *Critical Inquiry*, vol. 7, no. 3, 1981, pp. 577-603.

Shane, Tommy, et al. "From deepfakes to TikTok filters: How do you label AI content?" *First Draft*, 12 May 2021, <https://firstdraftnews.org:443/long-form-article/from-deepfakes-to-tiktok-filters-how-do-you-label-ai-content/>.

_____. "There are lots of ways to label AI content. But what are the risks?" *First Draft*, 13 May 2021, <https://firstdraftnews.org:443/long-form-article/there-are-lots-of-ways-to-label-ai-content-but-what-are-the-risks/>.

Skiena, Steven S. *The Algorithm Design Manual*. Springer International Publishing, 2020.

Sissons, Beth. "Freckles: Causes, Identification, and Risks." *Health News Today*, 26 Oct. 2018, <https://www.medicalnewstoday.com/articles/323471>.

Sturman, D. J. "Computer Puppetry." *IEEE Computer Graphics and Applications*, vol. 18, no. 1, 1998, pp. 38–45.

"Tackling the Misinformation Epidemic with 'In Event of Moon Disaster.'" *MIT News*, 20 July 2020, <https://news.mit.edu/2020/mit-tackles-misinformation-in-event-of-moon-disaster-0720>.

Toews, Rob. "Deepfakes Are Going To Wreak Havoc On Society. We Are Not Prepared." *Forbes*, 25 May 2020, <https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/>.

Vaccari, Cristian, and Andrew Chadwick. "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News." *Social Media + Society*, vol. 6, no. 1, SAGE Publications Ltd, Jan. 2020.

Vincent, James. “Tom Cruise Deepfake Creator Says Public Shouldn’t Be Worried about ‘One-Click Fakes.’” *The Verge*, 5 March 2021,

<https://www.theverge.com/2021/3/5/22314980/tom-cruise-deepfake-tiktok-videos-ai-impersonator-chris-ume-miles-fisher>.

Wahl-Jorgensen, Karin, and Matt Carlson. “Conjecturing Fearful Futures: Journalistic Discourses on Deepfakes.” *Journalism Practice*, vol. 15, no. 6, 2021, pp. 803–20.

Wardle, Claire. “This Video May Not Be Real.” *The New York Times Op Ed*, 14 Aug. 2019, <https://www.nytimes.com/video/opinion/100000006635241/deepfakes-adele-disinformation.html>.

Wilden, Anthony. *System and Structure: Essays in Communication and Exchange*, 2nd edition. Tavistock, 1980.

Wittgenstein, Ludwig. *On Certainty*. Edited by G.E.M. Anscombe and G.H. von Wright. Harper and Row, 1969.