Eye Contact with the Machine
Gaze Correction in Video Conferencing

Robert Rapoport and Vera Tollmann

During video-mediated communication it is not possible to look the person on screen in the eyes. While the problem is simple, the solution is not. Attempted remedies have variously been called *attention correction*, *gaze redirection*, or *gaze correction*. Gaze correction, on a basic level, seeks to fix the problem of a lens acting as a proxy for the eye. During a video call, one cannot simultaneously look at the screen and (the laptop/smartphone) lens, creating a dissonance: *Will we produce eye contact for ourselves or our conversation partner*? Since the pandemic's beginning, many have experienced the "uncorrected gaze" of video-mediated communication, metabolized by some as "Zoom fatigue" (see Distelmeyer and Lovink in this volume). The goal of gaze correction, succinctly stated, is "to digitally alter the appearance of eyes in a way that changes the apparent direction of the gaze" (Ganin et al. 2016). It could be argued that gaze correction, narrowly applied, simply fixes a mechanical problem: we have to look at a screen instead of a lens when in a video call. But eye contact is so foundational to communication, its meaning so tacit, that unintended consequences need to be considered.

Social cognition, our ability to make sense of the behavior of others, relies on both linguistic and nonlinguistic cues. The consensual imperative of unmediated eye contact stimulates the release of the neurotransmitter oxytocin. Studies of video conferencing have found that this phenomenon does not work as well on screen (Auyeung et al. 2015). Moreover, neuroscience has recently identified concrete physiological effects of live rather than prerecorded screen-mediated eye contact (Noah et al. 2020). Near-infrared spectroscopy found increased coherence in brain activity during *live* eye-to-eye contact. Additional studies have described the reciprocal prelinguistic processing of information as the "rapid and reciprocal exchange of salient information in which each send and receive 'volley' is altered in response to the previous [signal]" (Hirsch et al. 2017, 314). Thus, if eye contact is essentially *relational*, gaze correction in its current form removes this relationality, delegating the task of eye contact to a platform.[1] To build on the metaphor of the volley: gaze correction could be akin to watching a game of tennis in which every shot is "corrected" so as to be returnable. Such a game, while graceful for a time, would quickly become lifeless for viewer and spectator alike.

Gaze correction will present us with a choice: Do we perform the essential emotional labor of maintaining eye contact, or do we delegate it? If two people do consent to have a platform mediate eye contact for them, what metric will be used by the platform to decide when contact should be stopped and started? On a social level, will this produce a space in which we are constantly second-guessing eye contact? Might gaze correction create affect-specific platforms in which users can choose between styles of eye contact and thus attention (e.g., engaged, skeptical, or rapt)? This would continue the logic of "style transfer" (the practice of transposing qualitative metrics between images) that already exists on various social platforms (see figs. 7 and 8). When one is able to simply select different styles of eye contact, gaze *optimization* rather than *correction* would seem more apt.

Computational optimization involves increasing the efficiency of a system while decreasing its resource use (Sedgewick 1984, 84). The questions for machine learning-based gaze correction is what the primary object of optimization will be? If Zoom optimized for attention rather than fidelity to the user's intent, what kind of social space would result?

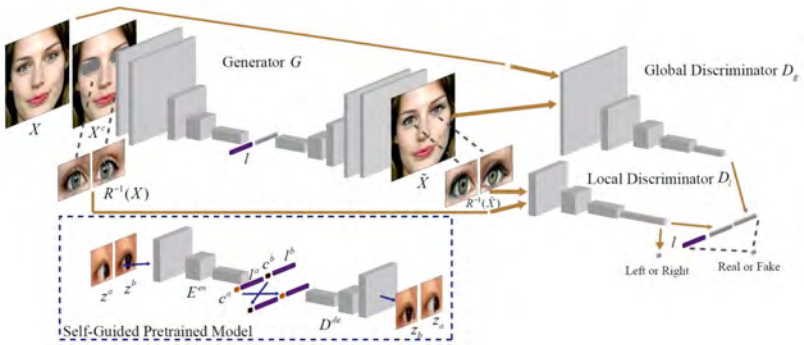
The state of the art in gaze correction involves using a generative adversarial network (GAN) to make real-time modifications to the user's eyes. A GAN consists of two neural networks: a classifier and a generator (fig. 1). The generator's role is to render a constant stream of training images because "the classification constraint does not work without the adversarial learning, or in other words, the adversarial learning helps to avoid adversarial examples" (He et al. 2020). The best-known use of GANs is the deepfake paper from 2016 (Thies et al. 2016). The epistemic threat of this technology has been widely discussed (Fallis 2020). To extrapolate: gaze correction is a real-time deepfake intervening at the most fundamental level of social communication, eye contact. While not as obviously threatening for public discussion as deepfakes, gaze correction nonetheless poses a similar epistemic threat (fig. 8). Jonathan Crary has written that the "reductiveness" of machine learning causes "splintering of the interhuman basis of a shared social reality" (Crary 2022, 92). Widespread gaze correction adoption would render it particularly difficult to track instances of such splintering precisely because we lack the vocabulary to talk about

1 At the time of writing, the authors were unable to find research on how this relational quality of eye contact would be encoded in gaze correction. We speculate that this is due to the lack of a training dataset. Such a dataset would only be produced if/when gaze correction were widely adopted. Widespread adoption of a gaze correction program without some mechanism to account for the consensuality of gaze could produce unintended effects.

such unprecedented mediation. General terms like "Zoom fatigue" are symptomatic of this lack of vocabulary. Bailenson (2021) has argued that video conferencing is tiring in part because our social brain expects a level of feedback that is not forthcoming. While this disconnect is a problem for our social cognition, a platform might see it as a way to build dependence. It is precisely those patterns of social behavior that *precede* language that are most attractive to those who seek to make us click before becoming aware of it. If we are engaging in conversations where the "person" looking at us is in fact a machine learning model, a whole set of ethical and epistemological questions follow. Again, the opposition between correction (social cognition) and optimization (machine learning) provides a shorthand to think through this shift. There is no ruleset for when to maintain eye contact, though if gaze correction is widely adopted there functionally will be, even if known only to the implementing platform.

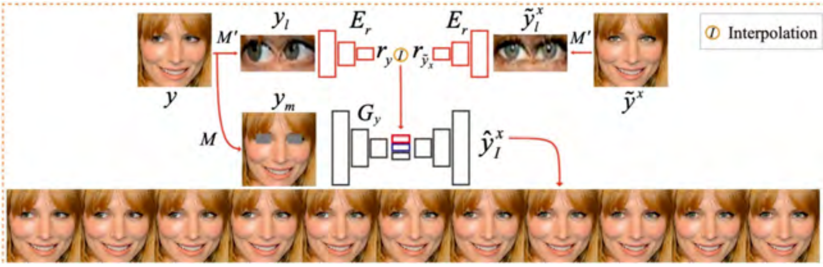
Gaze correction has been researched since the 1990s in both industry and academia. The most familiar attempt to roll out gaze correction was Apple's Eye Contact setting in iOS 13 (Bohannon 2013, 177). GPU producer Nvidia has attempted to take the technology a step further by redrawing not only eyes but also the interlocutor's head (Wang et. Al 2021). In what follows, we refer primarily to Zhang et al. (2020) as representative of the broad trend to correct gaze computationally rather than mechanically. The paper's GazeGAN model takes the initial live video feed and redraws the participant's pupils in real time (see fig. 4). This technique is called "Image Inpainting … an important task in computer vision and graphics aims to fill the missing pixels of an image with plausibly synthesized contents" (Zhang et al. 2019).

Figure 1: Overview of GazeGAN network architecture



Source: Zhang et al. 2019.

Figure 2: Architecture of GazeGAN "in-painting" model



Source: Zhang et al. 2020.

Zhang et al.'s focus is primarily on how realistic the redrawn ("in-painted") pupils and eyes are (fig. 2). This concern for realism, however, does not extend to the social realm. The GazeGAN model cannot recreate the cadence of mutual eye contact, which includes involuntary looking away and other subtleties of joint attention (Bohannon et al. 2012). Such relational interpersonal behavior is difficult to encode. Our decision to maintain eye contact depends on tacit knowledge of the other that we are often unaware of. What would such a training set look like? To date, there exists no training the authors could find. This is not to say that there could not be several diverse ones. Such training sets would have to index qualitative values like empathy, which produce believable, authentic eye contact in lived experience. In a way, video conferencing sessions with machine learning interventions could prevent harmful, unsettling behavior, such as contemptuous looks and eyes full of doubt.

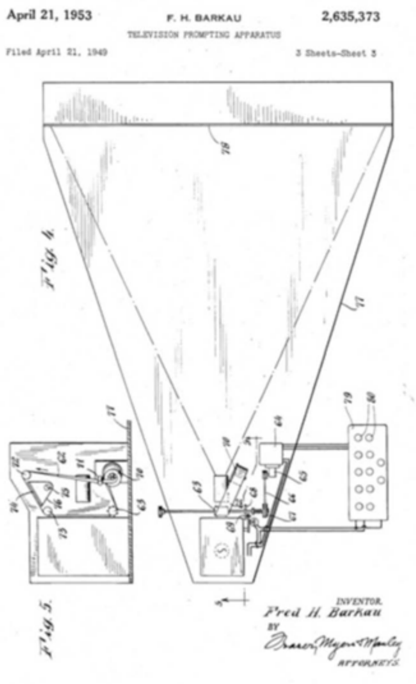
Hardware Antecedent to the *Split Gaze*

To understand the possible effects of gaze correction, we can look to one of its antecedents: the teleprompter, which shifts the eyes mechanically as opposed to computationally. In the early years of broadcasting, the teleprompter was developed as an aid for presenters of all types in the industry. In the postwar years, only professional film studios could afford the necessary hardware. Initially, the problems of the teleprompter were similar to those of twentieth-century video conferencing: *to align the display with the optical path of the camera*. The first "television prompting apparatus," patented in 1953, replaced a human prompter (who would read the text and give vocal cues when needed) (fig. 3).[2] The first generation of devices produced by the

2 Barkau, Fred H. 1953. Television Prompting Apparatus, US Patent 2,635,373A, filed April 21, 1949, and issued April 21, 1953.

TelePrompTer Corporation were attached to the side of the camera lens, making the presenter's gaze oblique to the viewers. Audiences at home could sense the presence of the mediating apparatus. After several iterations of the device, the shifted gaze of the presenter was effectively "corrected." In 1959, Jess Oppenheimer patented the first in-camera teleprompter, which used mirrors to project the script in front of the lens.[3] Below is a diagram from the initial patent.

Figure 3: Original patent diagram from Fred H. Barkau, 1953



Source: Fred H. Barkau, 1953. Television Prompting Apparatus, US Patent 2,635,373A.

The mechanical gaze correction of teleprompters thus allowed a presenter to appear to look the individual viewer in the eye, laying out the tension more completely realized in computational gaze correction: it simulates eye contact while allowing the speaker to read from a script. The news broadcast, for example, gained a new

3 Oppenheimer, J. 1959. Prompting Apparatus US Patent 2,883,902, filed Oct. 14, 1954, and issued April 28, 1959

intimacy of address through this technical trick. And yet their mediation simultaneously created a power imbalance: the corrected gaze of the broadcaster meant the news was internalized using, to some degree, social cognition. The teleprompter thus effaced a one-to-many flow of information. Might gaze correction create a similar dynamic for users of video conferencing applications? What happens in a situation in which every user can feign a direct gaze? What subtler prompts might this assemblage be capable of?

Neural Nets, Datasets, and Gaze Correction

The foundations for today's developments in gaze correction were laid in 2009 with the publication of ImageNet—the largest ever training data set for machine vision (Deng et al. 2009). Because of its size and granularity ImageNet greatly accelerated research in the field. In 2012, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) was won by a team from the University of Toronto using a neural network (Hinton et al. 2012). This watershed moment is considered the start of the widespread use of neural nets in machine learning research. So many papers used ImageNet for training models that inherent biases in how it labeled images created ripple effects throughout areas of society (see Crawford and Paglen 2019).

The other lynchpin in the development of contemporary gaze correction was the publishing of the Columbia Gaze Data Set in 2013 (see fig. 6). The data set provides a systematic collection of 5,880 images of fifty-six people over five head poses and twenty-one gaze directions. It marked the most comprehensive and diverse data set related to gaze correction to date, even though it was initially created at Columbia University for training purposes in human-object interaction, specifically, "to train a detector to sense eye contact in an image" (Smith et al. 2013, 271). This data set, coupled with the learning neural nets, has served to accelerate research on gaze correction. Indeed, the research by Zhang et al. (2020) discussed above relies on both. Prior to the Columbia Gaze Data Set, data sets were generally scraped from the web rather than purpose built. These data sets attempt to model a mode of communication that is not situated. In a sense, the corrected gaze appears to employ the long-criticized "view from nowhere." Dan Kotliar has pointed out that the process of dataset compilation often glosses over the fact that users exist between "very distant localities" (2020, 922). This same tendency is evident in the training sets for gaze correction.

Apart from the cultural and geographical context, on a data level, as Taina Bucher (2020) elaborates with Jean-Luc Nancy's concept of "being as always already a being-with," everyone is already existing in a connected state. By applying Nancy's concept to the social networks, she argues that besides active users, offline actors are also connected with others (i.e., as contacts). This way, even those who are not logged in leave data traces, since machine learning algorithms classify, recommend, and match according to data input. In a different publication, Bucher (2021, 100) refers to these passive data doubles as "non-users." In comparison to the algorithmic layers relevant for gaze correction, training data sets of eyes and faces, such as the Columbia Gaze Data Set, are the ground truth that connects any one pair of eyes with another. While participating in video conferences or casual group calls with eye contact enabled, users might eventually experience the state of "being singular plural" in a rarely visible way.

Social Cognition, Gaze Direction, and Video Conferencing

Eye contact is the infant's first step toward building social cognition. While following the gaze of our parents, we begin to understand how the world appears to the other (Frith 2010; Lamm et al. 2016). "The ability to process its [gaze] direction starts very early in life and is fundamental for the development of normal social cognition" (Itier and Batty 2009, 845). Some researchers have even argued that our theory of mind begins with gaze tracking (Farroni et al. 2002). Gaze direction—and the attention it engenders—is so tacit in our making of social meaning that it is difficult to reflect on, creating a "non-verbal interpersonal communication including salient and emotional information" (Hirsch et al. 2017). And yet in video conferencing, the uncorrected gaze appears without the cues we expect. "Zoom fatigue" has been attributed in part to the fact that the mediated gaze is "perceptually realistic, but not socially realistic" (Farroni et al. 2002). As this dissonance is prelinguistic, how will we metabolize its predictive "correction" by machine learning? Put differently, is there a deep patterning to eye contact that machine learning will be able to "see" in a way we cannot? The "cooperative eye hypothesis" argues that the particular morphology of the human eye—compared to other primates—facilitates higher levels of joint attention. The whites of human eyes enable us to read more intention from the eyes of others, facilitating cooperation. Having a gaze legible to others is thus intrinsic to our humanity (Tomasello et al. 2007).

Eye contact is rarely discussed in procedural terms, but machine learning-based approaches to gaze correction will require as much. At the front end, or "surface," these systems learn to redirect pupils in real time. "The screen is the surface, the display buffer is the subface of the algorithmic thing that the two of us—we ourselves and the program—are engaged in. The algorithmic thing comes as a visible appearance for us" (Nake 2015, 106). Gradually, turning on one's webcam may signal tacit acceptance of a new intimacy with machine learning. The delicacy of this moment is evident in the fits and starts with which the technology has been approached by companies as formidable as Apple. Apple's Eye Correction feature brought gaze correction to FaceTime in iOS 14 (Peterson 2020). First iterations were tested under the label of "attention correction." The feature's name was later changed to the more benign Eye Contact, perhaps in a bid to sound less disciplinary. A cursory search for this feature will surface many tutorials about disabling this "creepy" new feature. If Eye Contact returns to iOS—and it likely will—the questions in this chapter will become much more immediate.

To apply the language of computation to the social world, we might say that humans are evolutionarily optimized to read the gazes of others. Each culture has differing tacit norms around eye contact, all optimizing for different outcomes, stochastically. Would an iOS-based gaze model account for cultural specificity? As ever, this would be a hard system to reverse engineer. Thus, video conferencing applications equipped with eye tracking and gaze correction would become obvious places to study the connections between gaze and more complex behaviors that have been intuited but not modeled. Once a model is developed, it would be hard for platforms to resist manipulating eye contact for specific outcomes.[4] The human infant recruits caretaking behavior from its parents using gaze. Could a platform do the same, using gaze correction as a means of inducing bonding between users, and by extension with the platform itself? While our question is motivated by social concerns, companies in platform capitalism follow market rules and advances in AI. (For example, Zoom started using user content to train their AI in July 2023. Zoom offers a virtual assistant and AI generated content such as conference summaries).

Indeed, biometric signals are already creating new approaches in behavioral health as they can help predict illness before it appears clinically (Bednarik et al. 2005). This medical approach to eye tracking overlaps with the larger field in behavioral medicine called digital phenotyping. The technique argues that the data traces left on the touch screen (and the other sensors on the smartphone) can be correlated to neurological states (Insel 2017). Insel's article does not list eye tracking, but its capture could follow the same logic. One can imagine that a society built on video-mediated communication would quickly produce a granular model of gaze direction that could be correlated to any number of other biometrics—as digital phenotyping attempts to do with thumbs, voice, and speech. The long-term population-level effects of such a data set would be profound. A data set of eye movements during video conferences could yield many insights for digital phenotyping (Nag et al. 2020).

4 For example, in a video conference about contractual negotiations, a platform could "optimize" gaze correction to nudge both parties toward agreement. This is a fairly heavy-handed example, but the subtler behaviors that gaze correction could model are abundant.

Theories of the Gaze

In some cultures, the gaze is perceived as an agent in its own right, particularly in the archetype of the "evil eye" (Breuer 2015). The term, widespread since antiquity, describes the belief that a gaze can inflict harm. One can "cast an eye" with the intent to punish someone. De-potentialization and fluidity of "negative" and "positive" looks characterize the modern discussion of the evil eye; where tradition is abstracted, which is one of the prerequisites for its circulation (van Loyen 2015). In this sense, eye contact with the machine could literally mean being subject to the agency of machine vision, adding a new case to the splitting of eye and gaze, the matter of making visible and being seen in public, studied by Sartre as well as Goffman. At different times and with regard todifferent media, Barthes and, later, Silverman investigated the mediated gaze.

In a memorable scene, Sartre depicts momentarily forgetting one's own body in the immersion of the clandestine gaze (Sartre 1956, 277). A person squats in front of a keyhole to watch others inside a room. The sound of footsteps in the hallway makes him feel "suddenly struck in his being." At this moment, the observer sees himself all at once from the outside, as a voyeur. Sartre suggests: the one who is looked at becomes acutely aware of having a body. Here the gaze has an activating function. In video conferencing, by contrast, each participant is constantly self-aware due to the mirror-like self-video.

When gazed upon, we are objectified, aware that we are for the other an object of consciousness; this triggers a new, reflexive relationship to ourselves. For Sartre, the gaze is existential. The existence of those looked at manifests itself from another point of view. What does it mean, then, when in video conferencing this reciprocal balance is delegated to machines?

There is a crucial difference between the perception of Sartre's individual gaze as an agent in an analog space and the unified space of a corrected gaze that glosses over individual differences. We may become more cognizant of our individual habits once gaze correction throws them into relief. Following Sartre's train of thought, then, we can interpret the GAN as a proxy for the real-world situation described by his phenomenological approach. In the particular case of machine learning–based eye contact, the structures of perception—the spectator's, the camera's, and the GAN's gazes—are all being shifted into a kind of "proxy politics" (Steyerl 2014). Communicating using GAN-augmented eyes means inhabiting the proxy.

Erving Goffman compared social life in public space to a theater play. People present themselves on the basis of roles. In "On Face-Work: An Analysis of Ritual Elements of Social Interaction" (1955), Goffman studied how communication routines were shaped by the media. He defined face-work as "the positive social value a person effectively claims" (Goffman 1967, 5). Goffman continues: "One's own face and the face of others are constructs of the same order; it is the rules of the group and the definition of the situation which determine how much feeling one is to have for a face and how this feeling is to be distributed among the faces involved" (Goffman 1967, 6). What does video conferencing do to our ability to perform face-work on the social level? Goffman might draw a conclusion similar to Sartre's: gaze correction does not restore this order, which video-mediated communication disrupts.

Whereas both Sartre and Goffman were reading unmediated gazes *en plein air*, Barthes focused on the representation of gaze in photography. In Barthes's view, what defines the gaze is its excessiveness. In "Right in the Eyes" ([1977] 1991), he suggests that science interprets the gaze in three combinable ways: as information, as relationships, and finally as possession. Accordingly, Barthes differentiates three functions: an optical, a linguistic, and a haptic role of the gaze, making it an unsteady sign. Barthes describes a person who appears to be looking straight at the spectator or photographer. He elaborates: "In reality, the portrait looks at no one, and I know it; it looks only into the lens, that is, into another, enigmatic eye: the eye of truth" ([1977] 1991, 240). The camera, like the GAN, seeks to self-efface. The GAN performs elaborate artifice in the name of making the medium disappear. We do not look into a lens, into an apparatus, but at the screen displaying a conversation partner.

In *The Threshold of the Visible World* (1996), Kaja Silverman develops the concept of the "productive look," which can see through the political motivations in images. She takes Farocki's film *Images of the World and the Inscription of War* (1989) as a starting point to consider the camera's departure from human vision. Aerial photographs intensified the move toward quantifiable images: "From this vantage point, the invention of the camera represents less a moment of rupture with earlier visual technologies than the moment at which their implicit disjuncture from the eye becomes manifest" (Silverman 1996, 143). The quantifiable image thus brought us closer to a paradigm in which machines not only model but act on increasingly unstructured data. Kotliar describes "algorithms' ability to characterize, conceptualize, and affect users" (2020, 919). The algorithmic gaze is thus "a multifocal one—as a gaze that stems from a complex combination of diverse types of lenses" (934). And here, lenses are employed in the metaphorical sense as a new instance of the *split gaze*: looking at the screen and the camera at the same time. Returning to Barthes, the gaze here becomes more about automation than relationship; eye contact is reduced to the level of information, becoming a steady sign in the process. Machine learning functionally separates the eye from the gaze, thereby splitting bodies from behavior while producing a "surveillant assemblage" (Haggerty and Ericson 2000). While Deleuze and Guattari criticize the "cementing" of the human face through social coding, gaze correction appears as an automated update. The space opened up by the intensities that build up in the interpersonal sphere short-circuit.

More recently, researchers at Strelka Institute have coined the term "peak face" to describe the declining reliability of the face as a source of information (Abbott et al. 2021). In the case of video-mediated communication, what might move into the void left after peak face? Might vocal intonation or chat come to the fore?[5]

In *The Right to Look*, Mirzoeff claims that "the right to look is not about seeing. It begins at a personal level with the look into someone else's eyes to express friendship, solidarity, or love. That look must be mutual, each person inventing the other, or it fails. As such it is *unrepresentable*" (2011, 1). Mirzoeff treats visuality as a tool of power and "a discursive practice that has material effects" (2011, 3).[6] With gaze correction however, we witness a new take on visuality and power, in which algorithms become disciplinary apparatuses[7]—exercising cultural practices such as sorting and classifying and therefore changing human knowledge and social experience. Building his argument on a decolonial discourse, Mirzoeff concludes that "classifying, separating, and aestheticizing together form ... a 'complex of visuality'" (Mirzoeff 2011, 3–4). He continues: "The *right to look* claims autonomy from this authority, refuses to be segregated and spontaneously invents new forms" (2011, 3–4). The representation of the gaze in video conferencing could be read as a result of digital authority since it might determine the nonverbal communication of two participants or several participants in a group chat. In the physical world, eyelines establish a real space that cannot be recreated on a flat screen. In a different text, Mirzoeff sees the increased surveillance in the digital sphere as a state response to "the anxiety that imperial subjects might start to think and act in common. Because prior to all law, there is a relation between people. We move from the social to the individual" (2014, 228). Mirzoeff points out that in the digital sphere, the common is repressed in favor of control. Therefore, he argues based on neuroscientific findings about "mirror neurons" that relations between people cannot be represented, since *live* eye-contact is a condition.

The concept of machine vision doing things in the world instead of seeing is further elaborated by Luciana Parisi with the idea of "negative optics." In "Negative Optics in Vision Machines," Parisi explores how the automation of vision seems to "re-

5 Secondary signifiers such as voices—female voices in particular (Siegert and Niebuhr 2021)—are perceptually affected by speech signal compression, and the chat might be the last component where machine learning does not yet discriminate—as long as chatbots are not involved in the conversation (Nvidia, for example, integrates chatbots in their Conversational AI program). While, on the other hand, the industry argues for those modifications in terms of accessibility for disabled participants, it becomes evident once again that technology genders, surveils, and modulates behavior.

6 Relating to this approach, video conferencing could be considered as a democratizing tool, particularly in times of social distancing, since it enables mediated face-to-face contact (if not limited by digital divide).

7 Several media studies researchers have expressed their concern that algorithms exercise too influence over social realities (Beer 2009; Bucher 2012; Gillespie 2014; Kitchin and Dodge 2011).

define vision in terms of a mediatic function that does not rely on light" (2020, 1281). Furthermore, she speaks directly to the production process of GANs: "the invisible image of machines is part of alien epistemologies," training a model on generative artifice (1283). This moves us away from the binary of visible and invisible such that "image feedback is no longer assured by the interaction with the world" (1281). Thus, Parisi leaves the machinic eye tied to non-generative media behind, which was at its time of conceptualization bound to vision.[8] In Parisi's view, reflecting on GANs means reflecting on a new kind of "vision machine" that "generates an avalanche of inputs" divorced from the real world. A video conference dependent to some degree on the alien epistemology of the GAN may cast "eye contact" in a different light—beyond Western ocularcentrism.

An Augmented Social Reality?

Gaze correction is a subtle form of augmented reality, creating an interpretive ambiguity for the user. If eye contact to date could include only two people, how will we metabolize a third presence? In what is commonly called augmented reality, virtual objects are usually clearly distinguishable from the physical environment onto which they are projected. However, an inherent risk of AR technology is that "there may be people interested in misleading us by creating virtual objects which are disguised in the world scene as if they were real objects" (Ariso 2017, 7). Insofar as gaze correction requires the technique to efface itself, it creates an interpretive grey area around eye contact in general.

The augmented gaze is different from augmented reality applications we are familiar with. Augmented reality systems such as the ones studied by Tolentino (2019) and Wegenstein (2010, 29) in the context of TikTok and Snapchat face filters are self-referentially worn on the human body. Google Glasses also function in this self-referential mode with displays responding to the individual user's head movements (Mainzer 2017, 27). Gaze correction changes the reality a video conference participant looks at according to the movement of their partner. Therefore, this machine learning intervention is not interactive for the individual user and distinct from cosmetic augmentations (see fig. 5). The participant cannot parse whether the eye contact is mutual or machine based. Here, we could speak of a covert augmented reality.

8 In the context of drone operation, filmmaker Harun Farocki saw a transition from images representing the material world to what he called the "operational image"—one no longer representing an object but rather serving as a cue for action (Farocki 2004). Artist and writer Trevor Paglen builds on Harun Farocki's operational image for contemporary machine vision with the term "invisible images"—neural net operations based on image data never seen by humans, but nonetheless exercising agency (2016). For Paglen (2016), operations inside the black box are an "invisible world of machine-machine visual culture."

On both ends, this equivocal gaze itself begins to carry less information if we cannot trust the movement of the eyes. Elsewhere in this volume, Lovink reminds us that Zoom's centralized protocol is not the only one available to us. The reciprocity inherent in a peer-to-peer model might metabolize gaze correction without the optimization. In either case, the drive of automated eye contact is to make itself indistinguishable from reality.

Conclusion: Gaze Correction versus Optimization

We have yet to see widespread adoption of gaze correction. If and when this happens, we will each have to decide if we turn it off.[9] What type of user might willingly delegate the maintenance of eye contact to machine learning? Or will the technology insinuate itself as a default? In the emerging discourse on the effects of online learning, a deficit in social learning is frequently cited (Strouse and Samson 2021). The interruption of social cognition created by video-mediated communication will, one can extrapolate, encompass gaze correction. But once the technology is implemented, it would be a matter of turning the dial on a parameter to socialize increasingly reward-based behavior. In the race down the brain stem, we know that some signals in eye contact are functionally preconscious. Thus, to engage in speculation: if the technology is socially assimilated on a given platform, it would be very tempting for that platform to begin optimizing for user retention over the social accuracy of the technique. In the language of industry, this would mean "favoring engagement" over other factors. What happens when two people are having a conference and have fundamentally different experiences of each other than face-to-face because of neural nets optimizing their gaze style? What would be the larger atomizing effects on society?

Were it to be narrowly applied, gaze correction could be a net benefit. Given the power of gaze to alter behavior there are clear multiple profit incentives for platforms to optimize interactions. A user would not have to do anything to appear social but accept the *correction*. However, as we have argued, true correction is a product of social cognition. GANs, by contrast, are systems composed of deterministic functions to favor an output. If that output is eye contact, how much is too much?

Mutual eye contact creates an inherently affective space. The tone of this space has to date resisted capture or encoding. During eye contact there are patterns at play that we *cannot* know and that a platform will nonetheless approximate. This behavioral data is functionally preconscious, placing the results outside of social discourse. The capture of eye contact patterns would thus appear a further step in

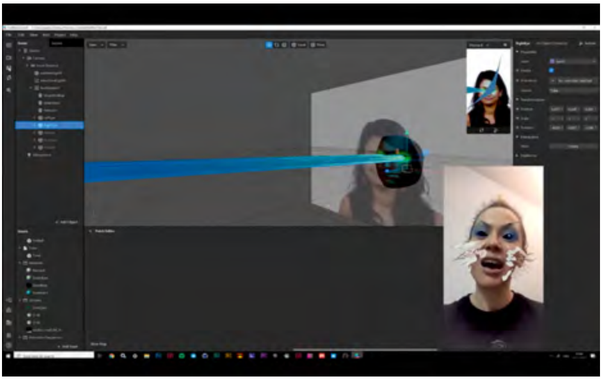
9 At the time of writing a cursory search for "ios, facetime, eye contact" reveals many want to disable it.

the black-boxing of the social world by proprietary code. Apple's Eye Contact will not be the last attempt to introduce gaze correction as a default. The question moving forward will be whether we will learn to recognize the change before it becomes functionally invisible. Recall the initial reaction to gaze correction: users googled the iOS 13's Eye Contact in order to learn how to disable it. Returning to Mirzoeff's reflection on the gradual loss of the commons, gaze correction can be seen as one such incursion. In this liminal space we should remember to ask who or what we are making eye contact with.

Figure 4: Qualitative comparison of gaze correction models

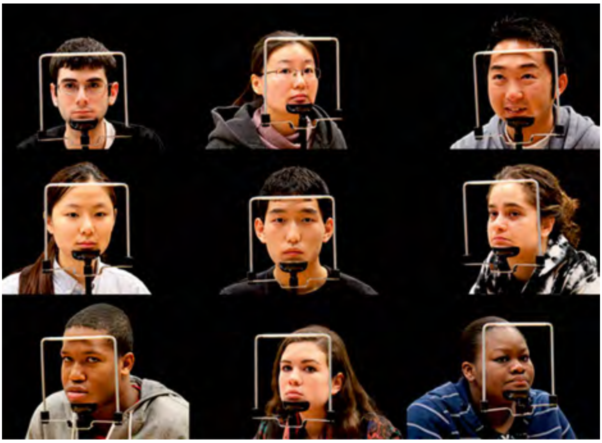
Source: Zhang et al. 2020.

Figure 5: An artistic response to machine vision and the automated gaze



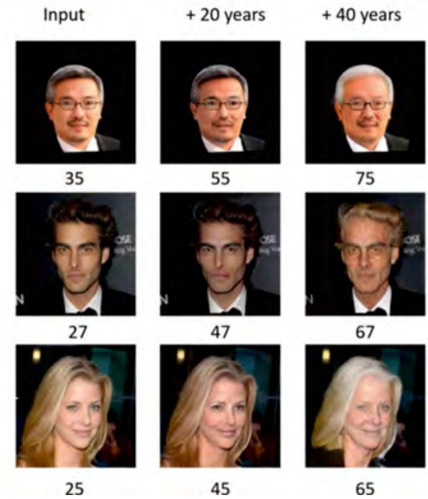
Source: Screenshot from Popp, Alla. 2021. "#alieneffect Face-filter Workshop for Beginners. Spy on Me #2 Online Program." HAU—Hebbel am Ufer, Berlin, accessed August 11, 2022. https://www.youtube.com/watch?v=lkp6zkyYh8Y.

Figure 6: Columbia Gaze Data Set



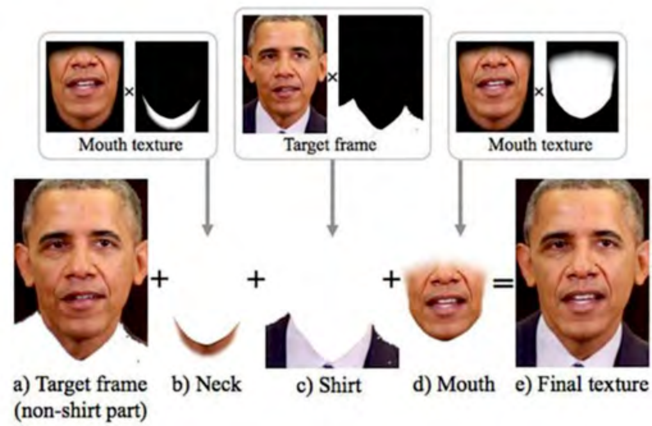
Source: Smith et al. 2013.

Figure 7: Generative adversarial style transfer networks for face aging



Source: Palsson et al. 2018.

Figure 8: An example of a deepfake synthesis procedure



Source: Suwajanakorn et al. 2017.

References

Abbott, M. C., Maria Buey González, and Carl Olsson (Strelka Institute). 2021. "Peak Face." YouTube video, 11:09, August 23, 2021. https://www.youtube.com/watch?v=rcfp25LaWjI.

Ahuja, Abhimanyu S., Vineet Pasam Reddy, and Oge Marques. 2020. "Artificial Intelligence and COVID-19: A Multidisciplinary Approach." *Integrative Medicine Research* 9 (3): 100434. https://doi.org/10.1016/j.imr.2020.100434.

Akechi, Hironori, Atsushi Senju, Helen Uibo, Yukiko Kikuchi, Toshikazu Hasegawa, and Jari K. Hietanen. 2013. "Attention to Eye Contact in the West and East: Autonomic Responses and Evaluative Ratings." *PLoS ONE* 8 (3): e59312. https://doi.org/10.1371/journal.pone.0059312.

Althusser, Louis. 2012. "Ideology and Ideological State Apparatuses." In *Media and Cultural Studies: Keyworks*, edited by Meenakshi Gigi Durham and Douglas M. Kellner, 200–4. Malden, MA: Blackwell.

Ariso, José María. 2017. "Is Critical Thinking Particularly Necessary when Using Augmented Reality in Knowledge Society? An Introductory Paradox." In *Augmented Reality*, edited by José María Ariso, 3–22. Berlin: De Gruyter. http://doi.org/10.26530/oapen_628401

Ashwin, Chris, and Paola Ricciardelli. 2019. "Editorial: Typical and Atypical Processing of Gaze." *Frontiers in Psychology* 10. https://doi.org/10.3389/fpsyg.2019.02576.

Bailenson, Jeremy N. 2021. "Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue." *Technology, Mind, and Behavior* 2 (1). https://doi.org/10.1037/tmb0000030.

Barthes, Roland. 1991. "Right in the Eyes." In *The Responsibility of Forms: Critical Essays on Music, Art, and Representation*, translated by Richard Howard, 237–242. Berkeley: University of California Press.

Barthes, Roland. 2015. "Auge in Auge." In *Roland Barthes: Auge in Auge*, edited by Peter Geimer and Bernd Stiegler, translated by Dieter Horning, 200–4. Berlin: Suhrkamp.

Bednarik, Roman, Tomi Kinnunen, Andrei Mihaila, and Pasi Fränti. 2005. "Eye-Movements as a Biometric." In *Image Analysis*, edited by Heikki Kalviainen, Jussi Parkkinen, and Arto Kaarna, 780–89. Berlin: Springer. https://doi.org/10.1007/11499145_79.

Bohannon, Leanne S., Andrew M. Herbert, Jeff B. Pelz, and Esa Rantanen. 2013. "Eye Contact and Video-Mediated Communication: A Review." Displays 34 (2): 177–85.

Breuer, Christian. 2015. *"Und wo blicke können tödten …": Der Malocchio als begehrlicher Augen/Blick*. Berlin: Kadmos.

Bucher, Taina. 2020. "Nothing to Disconnect from? Being Singular Plural in an Age of Machine Learning." *Media, Culture & Society* 42 (4): 610–17. https://doi.org/10.1177/0163443720914028.

Bucher, Taina. 2021. *Facebook*. Cambridge, UK: Polity.

Crary, Jonathan. 2022. *Scorched Earth: Beyond the Digital Age to a Post-Capitalist World*. New York: Verso.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "Imagenet: A Large-Scale Hierarchical Image Database." *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 248–55.

Ewing, Louise, Gillian Rhodes, and Elizabeth Pellicano. 2010. "Have You Got the Look? Gaze Direction Affects Judgements of Facial Attractiveness." *Visual Cognition* 18 (3): 321–30. https://doi.org/10.1080/13506280902965599.

Fallis, Don. 2020. "The Epistemic Threat of Deepfakes." *Philosophy & Technology* 1 (21): 623–43.

Farocki, Harun. 2004. "Phantom Images." *Public* 2 9: 12–22. http s://public.journals.yorku.ca/index.php/public/article/view/30354.

Farroni, Teresa, Gergely Csibra, Francesca Simion, and Mark H. Johnson. 2002. "Eye Contact Detection in Humans from Birth." *Proceedings of the National Academy of Sciences* 99 (14): 9602–5. https://www.pnas.org/doi/full/10.1073/pnas.152159999.

Fiske, Susan T., and Shelley E. Taylor. 2008. *Social Cognition: From Brains to Culture*. New York: McGraw-Hill.

Frith, Uta, and Chris Frith. 2010. "The Social Brain: Allowing Humans to Boldly Go Where No Other Species Has Been." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1537): 165–76.

Ganin, Yaroslav, Daniil Kononenko, Diana Sungatullina, Victor Lempitsky. 2016. "DeepWarp: Photorealistic Image Resynthesis for Gaze Manipulation." In *Computer Vision—ECCV 2016*, edited by Bastian Leibe, Jiri Matas, Nicu Sebe, Max Welling, 311–26. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-46475-6_20.

Glissant, Édouard. 1997. *Poetics of Relation*. Ann Arbor: University of Michigan Press.

Goffman, Erving. 1955. "On Face-Work: An Analysis of Ritual Elements." *Psychiatry* 18 (3): 213–31. https://doi.org/10.1080/00332747.1955.11023008.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. "Generative Adversarial Nets." *Advances in Neural Information Processing Systems 27* (NIPS 2014): 2672–80.

Haensel, Jennifer X., Tim J. Smith, and Atsushi Senju. 2021. "Cultural Differences in Mutual Gaze during Face-to-Face Interactions: A Dual Head-Mounted Eye-Tracking Study." *Visual Cognition* 30 (1–2): 100–15. https://doi.org/10.1080/13506285.2021.1928354.

Haggerty, Kevin D., Ericson, Richard V. 2000. "The Surveillant Assemblage." *The British Journal of Sociology* 51: 605–22.

He, Zhenliang, Wangmeng Zuo, Meina Kan, Shiguang Shan, and Xilin Chen. 2019. "AttGAN: Facial Attribute Editing by Only Changing What You Want." *IEEE Transactions on Image Processing* 28 (11): 5464–78.

He, Zhenyi, Keru Wang, Brandon Yushan Feng, Ruofei Du, and Ken Perlin. 2021. "GazeChat: Enhancing Virtual Conferences with Gaze Awareness and Interactive 3D Photos." *UIST '21: ACM Symposium on User Interface Software and Technology*. https://www.youtube.com/watch?v=Fnl8hxxgGVQ.

Hess, Ursula, and Agneta Fischer. 2013. "Emotional Mimicry as Social Regulation." *Personality and Social Psychology Review* 17 (2): 142–57. https://doi.org/10.1177/1088868312472607.

Hinton, Geoffrey E., Li Deng, Dong Yu, George E. Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N. Sainath, and Brian Kingsbury. 2012. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups." *IEEE Signal Processing Magazine* 29 (6): 82–97.

Hirsch, Joy, Xian Zhang, J. Adam Noah, and Yumie Ono. 2017. "Frontal Temporal and Parietal Systems Synchronize within and across Brains During Live Eye-to-Eye Contact." *NeuroImage* 157: 314–30. https://doi.org/10.1016/j.neuroimage.2017.06.018.

Insel, Thomas R. 2017. "Digital Phenotyping: Technology for a New Science of Behavior." *JAMA* 318 (13): 1215–16. https://doi.org/10.1001/jama.2017.11295.

Itier, Roxane J., and Magali Batty. 2009. "Neural Bases of Eye and Gaze Processing: The Core of Social Cognition." *Neuroscience & Biobehavioral Reviews* 33 (6): 843–63. https://doi.org/10.1016/j.neubiorev.2009.02.004.

Kiilavuori, Helena, Veikko Sariola, Mikko J. Peltola, and Jari K. Hietanen. 2021. "Making Eye Contact with a Robot: Psychophysiological Responses to Eye Contact with a Human and with a Humanoid Robot." *Biological Psychology* 158: 107989. https://doi.org/10.1016/j.biopsycho.2020.107989.

Kobayashi, Hiromi, and Shiro Kohshima. 2001. "Unique Morphology of the Human Eye and Its Adaptive Meaning: Comparative Studies on External Morphology of the Primate Eye." *Journal of Human Evolution* 40 (5): 419–35. https://doi.org/10.1006/jhev.2001.0468.

Kotliar, Dan M. 2020. "Data Orientalism: On the Algorithmic Construction of the Non-Western Other." *Theory and Society* 49 (5–6): 919–39. https://doi.org/10.1007/s11186-020-09404-2.

Kuang, Beibei, Xueting Li, Xintong Li, Mingxiao Lin, Shanrou Liu, and Ping Hu. 2021. "The Effect of Eye Gaze Direction on Emotional Mimicry: A Multimodal

Study with Electromyography and Electroencephalography." *NeuroImage* 226: 117604. https://doi.org/10.1016/j.neuroimage.2020.117604.

Lamm, Claus, Henryk Bukowski, and Giorgia Silani. 2016. "From Shared to Distinct Self–Other Representations in Empathy: Evidence from Neurotypical Function and Socio-cognitive Disorders." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1686): 20150083. https://doi.org/10.1098/rstb.2015.0083.

Liebling, Daniel J., and Sören Preibusch. 2014. "Privacy Considerations for a Pervasive Eye Tracking World." *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 1169–77. Seattle: ACM. https://doi.org/10.1145/2638728.2641688.

Madipakkam, Apoorva Rajiv, and Marcus Rothkirch. 2019. "The Unconscious Processing of Social Information." In *Transitions between Consciousness and Unconsciousness*, edited by Guido Hesselmann, 92–117. New York: Routledge. https://doi.org/10.4324/9780429469688-4.

Mainzer, Klaus. 2017. "From Augmented Reality to the Internet of Things: Paradigm Shifts in Digital Innovation Dynamics." In *Augmented Reality: Reflections on Its Contribution to Knowledge Formation*, edited by José María Ariso, 25–40. Berlin: De Gruyter. https://doi.org/10.1515/9783110497656-002.

Mirzoeff, Nicholas. 2014. "Visualizing the Anthropocene." *Public Culture*, 26 (2): 213–32.

Mirzoeff, Nicholas. 2011. *The Right to Look*. Durham, NC: Duke University Press.

Mirzoeff, Nicholas. 2002. "The Subject of Visual Culture." In *The Visual Culture Reader*, edited by Nicholas Mirzoeff, 3–23. New York: Routledge.

Müller, Philipp, Michael Dietz, Dominik Schiller, Dominike Thomas, Guanhua Zhang, Patrick Gebhard, Elisabeth André, and Andreas Bulling. 2021. "MultiMediate: Multi-modal Group Behaviour Analysis for Artificial Mediation." *Proceedings of the 29th ACM International Conference on Multimedia*, 4878–82. New York: Association for Computing Machinery. https://doi.org/10.1145/3474085.3479219.

Nag, Anish, Nick Haber, Catalin Voss, Serena Tamura, Jena Daniels, Jeffrey Ma, Br yan Chiang, Shasta Ramachandran, Jessey Schwartz, Terry Winograd, Carl Fei nstein, and Dennis P. Wall. 2020. "Toward Continuous Social Phenotyping: Analyzing Gaze Patterns in an Emotion Recognition Task for Children with Autism through Wearable Smart Glasses." *Journal of Medical Internet Research* 22 (4). https ://doi.org/10.2196/13810.

Nake, Frieder. 2015. "Surface, Interface, Subface: Three Cases of Interaction and One Concept." In *Paradoxes of Interactivity: Perspectives for Media Theory, Human-Computer Interaction, and Artistic Investigations*, edited by Uwe Seifert, Jin Hyun Kim, and Anthony Moore, 92–109. Bielefeld: transcript.

Nancy, Jean-Luc. 2000. *Being Singular Plural*. Redwood City, CA: Stanford University Press.

Niedźwiecka, Alicja. 2020. "Look Me in the Eyes: Mechanisms Underlying the Eye Contact Effect." *Child Development Perspectives* 14 (2): 78–82. https://doi.org/10.1111/cdep.12361.

Noah, J Adam, Xian Zhang, Swethasri Dravida, Yumie Ono, Adam Naples, James C. McPortland, and Joy Hirsch. 2020. "Real-Time Eye-to-Eye Contact Is Associated with Cross-Brain Neural Coupling in Angular Gyrus." *Frontiers in Human Neuroscience* 14 (19). https://doi.org/10.3389/fnhum.2020.00019.

Nomura, Tatsuya, Takayuki Kanda, Tomohiro Suzuki, and Sachie Yamada. 2020. "Do People with Social Anxiety Feel Anxious about Interacting with a Robot?" *AI & SOCIETY* 35 (2): 381–90. https://doi.org/10.1007/s00146-019-00889-9.

Oppenheimer, J. Robert. 1954. Prompting Apparatus, Patent Publication US2883902A. https://patents.google.com/patent/US2883902A/ (accessed August 18, 2022).

Palsson, Sveinn, Eirikur Agustsson, Radu Timofte, and Luc Van Gool. 2018. "Generative Adversarial Style Transfer Networks for Face Aging." In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2165–68. Salt Lake City, UT: IEEE. https://doi.org/10.1109/CVPRW.2018.00282.

Paripally, Gopal, Craig Clapp, and Tony Pham. 2012. Reflective and Refractive Solutions to Providing Direct Eye Contact Videoconferencing. US Patent 8,970,655, filed December 14, 2012, and issued March 3, 2015. https://patents.google.com/patent/US8970655B2/en?q=teleprompter+mirror&oq=teleprompter+mirror+a (accessed August 18, 2022).

Parisi, Luciana. 2020. "Negative Optics in Vision Machines." *AI & Society* 36: 1281–93. https://doi.org/10.1007/s00146-020-01096-7.

Pasini, Marco. 2019. "Style Transfer with GANs on HD Images." *Medium*, July 23, 2019. https://towardsdatascience.com/style-transfer-with-gans-on-hd-images-88e8efcf3716.

Pathirana, Primesh, Shashimal Senaratha, Dulani Meedeniya, and Sampath Jayarathna. 2022. "Eye Gaze Estimation: A Survey on Deep Learning-Based Approaches." *Expert Systems with Applications* 199: 116894.

Peterson, Mike. 2020. "FaceTime Eye Contact Correction Feature to Launch with IOS 14." Apple Insider, June 22, 2020. https://appleinsider.com/articles/20/06/22/facetime-eye-contact-correction-feature-to-launch-with-ios-14.

Popp, Alla. 2020. "#alieneffect Facefilter Workshop für Anfänger*innen. Spy on Me #2 Online Program." YouTube video, March 20, 2020. https://www.youtube.com/watch?v=lkp6zkyYh8Y.

Sartre, Jean Paul. 1956. *Being and Nothingness: An Essay on Phenomenological Ontology*. Translated by Hazel E. Barnes. New York: Philosophical Library.

Schukin, Dave. (@schukin). 2019. "How iOS 13 FaceTime Attention Correction works: it simply uses ARKit to grab a depth map/position of your face, and adjusts the eyes accordingly. Notice the warping of the line across both the eyes and nose."

Twitter, July 3, 2019, 12:07 p.m. https://twitter.com/schukin/status/1146359923158089728.

Sedgewick, Robert. 1984. *Algorithms*. Reading, MA: Addison-Wesley.

Senju, Atsushi, and Mark H. Johnson. 2009. "The Eye Contact Effect: Mechanisms and Development." *Trends in Cognitive Sciences* 13 (3): 127–34. https://doi.org/10.1016/j.tics.2008.11.009.

Shimojo, Shinsuke, Claudiu Simion, Eiko Shimojo, and Christian Scheier. 2003. "Gaze Bias Both Reflects and Influences Preference." *Nature Neuroscience* 6 (12): 1317–22. https://doi.org/10.1038/nn1150.

Siegert, Ingo, and Oliver Niebuhr. 2021. "Case Report: Women, Be Aware That Your Vocal Charisma Can Dwindle in Remote Meetings." *Frontiers in Communication* 5. https://doi.org/10.3389/fcomm.2020.611555.

Silverman, Kaja. 1996. *The Threshold of the Visible World*. New York: Routledge.

Smith, Brian A., Qi Yin, Steven K. Feiner, and Shree K. Nayar. 2013. "Gaze Locking: Passive Eye Contact Detection for Human–Object Interaction." *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, 271–280. UIST '13. https://doi.org/10.1145/2501988.2501994.

Srnicek, Nick. 2017. *Platform Capitalism*. New York: John Wiley & Sons.

Steyerl, Hito. 2014. "Proxy Politics. Signal and Noise." *e-flux journal* 60. https://www.e-flux.com/journal/60/61045/proxy-politics-signal-and-noise.

Strouse, Gabrielle A., and Jennifer E. Samson. 2021. "Learning from Video: A Meta-Analysis of the Video Deficit in Children Ages 0 to 6 Years." *Child Development* 92 (1): e20–38. https://doi.org/10.1111/cdev.13429.

Suwajanakorn, Supasorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman (2017). "Synthesizing Obama: Learning Lip Sync from Audio." *ACM Transactions on Graphics (ToG)* 36 (4): 1–13.

Thies, Justus, Michael Zollhöfer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. 2016. "Face2Face: Real-Time Face Capture and Reenactment of RGB Videos." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 2387–95.

Tolentino, Jia. 2019. "The Age of Instagram Face." *The New Yorker*, December 12, 2019. https://www.newyorker.com/culture/decade-in-review/the-age-of-instagram-face.

Tomasello, Michael, Brian Hare, Hagen Lehmann, and Josep Call. 2007. "Reliance on Head versus Eyes in the Gaze Following of Great Apes and Human Infants: The Cooperative Eye Hypothesis." *Journal of Human Evolution* 52 (3): 314–20. https://doi.org/10.1016/j.jhevol.2006.10.001.

van Loyen, Ulrich. 2015. "Herzensergießungen. Süditaliens böser Blick." *Zeitschrift für Kulturwissenschaften. Begeisterung und Blasphemie* 9 (2): 89–95. https://doi.org/10.25969/mediarep/13940.

Vidal, Mélodie, Jayson Turner, Andreas Bulling, and Hans Gellersen. 2012. "Wearable Eye Tracking for Mental Health Monitoring." *Computer Communications* 35 (11): 1306–11. https://doi.org/10.1016/j.comcom.2011.11.002.

Wang, Ting-Chun, Arun Mallya, Ming-Yu Liu. 2021. "One-Shot Free-View Neural Talking-Head Synthesis for Video Conferencing." *CVPR*. https://nvlabs.github.io/face-vid2vid.

Wegenstein, Bernadette. 2010. "Body." In *Critical Terms for Media Studies*, edited by Mark B. N. Hansen and W. J. T. Mitchell, 19–34. Chicago: University of Chicago Press.

Wood, Erroll, Tadas Baltrusaitis, Louis-Philippe Morency, Peter Robinson, and Andreas Bulling. 2017. "GazeDirector: Fully Articulated Eye Gaze Redirection in Video." *arXiv:1704.08763 [cs.CV]*. http://arxiv.org/abs/1704.08763.

Zhang, Jichao, Meng Sun, Jingjing Chen, Hao Tang, Yan Yan, Xueying Qin, and Nicu Sebe. 2019. "GazeCorrection:Self-Guided Eye Manipulation in the Wild Using Self-Supervised Generative Adversarial Networks." *arXiv:1906.00805 [cs.CV]*. https://arxiv.org/abs/1906.00805.

Zhang, Jichao, Jingjing Chen, Hao Tang, Wei Wang, Yan Yan, Enver Sangineto, and Nicu Sebe. 2020. "Dual In-Painting Model for Unsupervised Gaze Correction and Animation in the Wild." *Proceedings of the 28th ACM International Conference on Multimedia*, 1588–96. Seattle: ACM. https://doi.org/10.1145/3394171.3413981.