



Introduction: ways of machine seeing

Mitra Azar¹ · Geoff Cox² · Leonardo Impett³

Received: 23 November 2020 / Accepted: 26 November 2020 / Published online: 20 February 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd. part of Springer Nature 2021

How do machines, and, in particular, computational technologies, change the way we see the world? This special issue brings together researchers from a wide range of disciplines to explore the entanglement of machines and their ways of seeing from new critical perspectives. As the title makes clear, we take our point of departure in John Berger's 1972 BBC documentary series *Ways of Seeing*, a four-part television series of 30-min films created by Berger and producer Mike Dibb, which had an enormous impact on both popular and academic perspectives on visual culture. Berger's scripts were adapted into a book of the same name, published by Penguin also in 1972. The book consists of seven numbered essays: four using words and images; and three essays using only images. Seeing is evidently a political act, exemplified in the third episode-chapter, where images of women in early modern European painting (Pol de Limbourg, Cranach the Elder, Jan Gossaert, Tintoretto) and commercial magazines are juxtaposed to demonstrate the ways in which women are rendered as objects of the male gaze. More broadly, Berger emphasised that "the relation between what we see and what we know is never settled". In this special issue, we explore how these ideas can be understood in the light of technical developments in machine vision and algorithmic learning, and how the relations between what we see and know are further unsettled.

What we see above (Fig. 1) is clearly not a book as such but a technically reproduced image of a book (Cox 2016). This testifies to the ways in which what, and how, we see and know is further unsettled through complex assemblages of elements, the details of which are largely kept from view. Access to the means of production here emphasises the Marxist approach of Berger (and, in turn, the Marxism of Benjamin, who Berger credits with furnishing the key concepts in *Ways of Seeing*), and how exposing what stays invisible allows a political understanding of social relations and thereby the possibility of their transformation. The TV programmes employed Brechtian techniques: revealing the technical apparatus of the studio, to encourage viewers not to simply watch (or read) in a straightforward way but rather be forced into an analysis of their alienation. Although rather less sophisticated in form, we hope for something similar with this journal that readers will reflect on what they see and read in ways that lead to a "return from alienation" or recognition of distancing-effects—breaking the "fourth wall" of machine vision, so to speak—to expose the unevenness of social relations in new ways.

Alienated forms of social interaction have become the "new normal" as we write. The current pandemic seems to heighten uncertainties about what is rendered visible and invisible to human perception. When the visual field is increasingly nonhuman, how is the world made knowable to us when much of its operations lay outside our visual register and consequently outside the scope of human action? Adrian Mackenzie and Anna Munster refer to an "operationalization" of visibility, in which images operate within a field of "distributed invisibility" in which relations between images count more than indexicality or iconicity (we might add, "aura" (Benjamin 2008)) of a single image (Mackenzie and Munster 2019: 16). Seeing, or what they call "platform seeing", becomes distributed through data practices and machinic assemblages that "emphasize the importance of the formatting of image ensembles as datasets across contemporary data practices; the incorporation of platforms into hardware in devices; forms of parallel computation; and the computational architectures of contemporary artificial

✉ Geoff Cox
geoffcox@lsbu.ac.uk

Mitra Azar
azar@cc.au.dk

Leonardo Impett
leonardo.l.impett@durham.ac.uk

¹ Department of Digital Design and Information Studies, Aarhus University, Aarhus, Denmark

² Centre for the Study of the Networked Image, London South Bank University, London, UK

³ Department of Computer Science, Durham University, Durham, UK

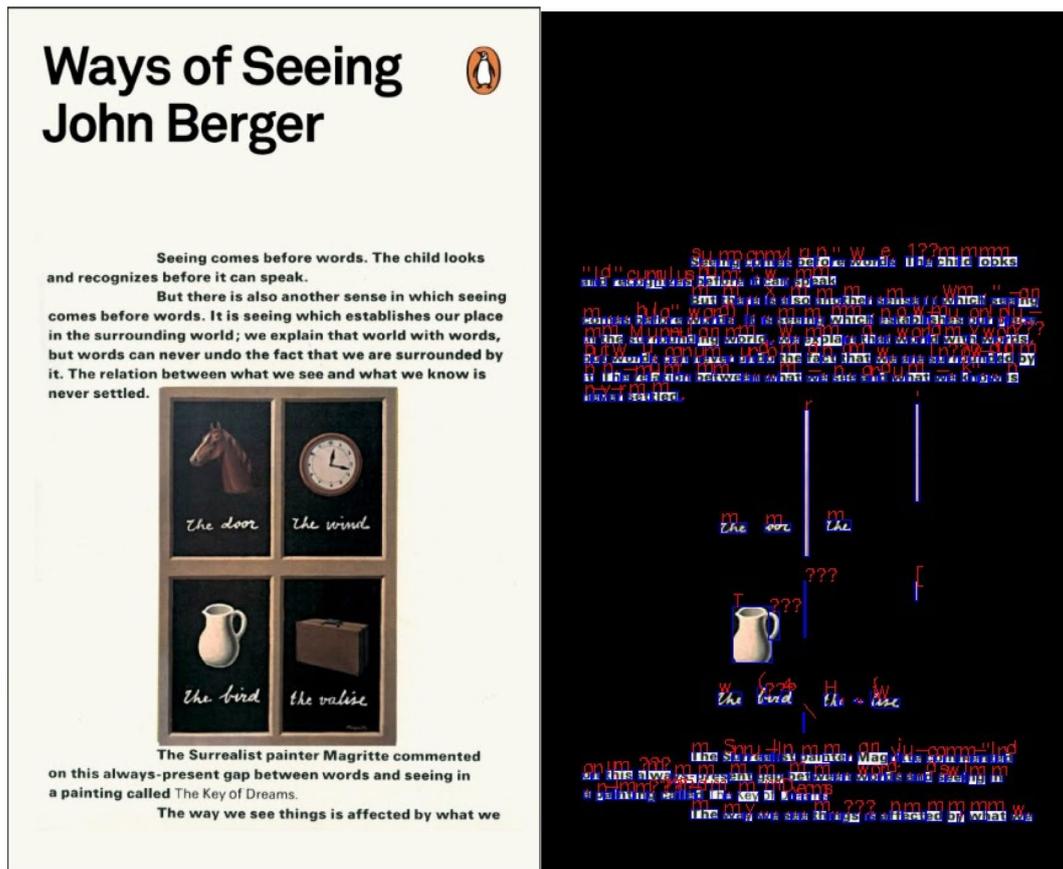


Fig. 1 The Cover of *Ways of Seeing* by John Berger (1972), and as seen through an optical character recognition programme. Images from Penguin Books and Scandinavian Institute for Computational Vandalism

intelligence. These assemblages constitute the (nonhuman) activities of perception as mode of cutting into/selecting out of the entire flux of image-ensemble world” (2019: 3). The importance of this understanding is that a new mode of perception is operationalized, what they call “invisual perception” (2019: 4): a new way of (machine) seeing which is an assemblage of its various parts; including imaging devices (such as cameras), the data they produce (which might take the form of an image), and the wider practices and infrastructures through which they are operationalized (in terms of its application).

So how to begin to understand the relation between seeing and knowing within this new operational context? Artist Trevor Paglan bluntly explains: “We’ve long known that images can kill. What’s new is that nowadays, they have their fingers on the trigger” (2014). This intensification of visuality takes “necropolitical” form (defined by Achille Mbembé (2003) as the question of who lives and who dies). Central to this is the “right to look” as Nicholas Mirzoeff has previously put it, the need to reclaim autonomy from authority, and generate new forms of “countervisuality” that turn the unreality created by visuality’s fake authority into real alternatives

(2011: 476, 485). For this journal, the special relationship between the formation of “racial surveillance capitalism” and the artificial vision of colonial power is further elaborated by Mirzoeff, for instance, as a form of fake authority over asylum seekers and refugees (see “Artificial Vision, White Space and Racial Surveillance Capitalism”, herein). The “realism” of this authority is rendered open to question, and some of the contradictions inherent to machine seeing are exposed. Indeed, “Reality is not only everything which is, but everything which is becoming. It’s a process. It proceeds in contradictions. If it is not perceived in its contradictory nature it is not perceived at all” (Brecht, cited in Mirzoeff 2011: 477).

We might well ask how machine seeing further exposes contradictions, or perhaps there is another kind of negation at work? To paraphrase, Luciana Parisi in “Negative Optics in Vision Machines” (herein), can machine vision step beyond the “oculocentric metaphysics of the Western gaze and the reproduction of racial capital”? She refers to this “inhuman” mode of machine vision as “negative optics” (like countervisuality) in order to offer an internal critique of ocular metaphysics, but also to defy the equation of value

between 0 and 1 s that sustains the universal law of (computational) capital. By starting from the negativity of the image, the racialized and gendered conditions of “artificial intelligence capitalism” demonstrate that “the equation of value maintains the condition of the zero of blackness”. What is challengingly proposed is that randomness in computation is part of the expansion of “heretic epistemologies” that start from dark optics and address what Denise Ferreira Da Silva calls “blackness”, namely “matter without form”, or “matter beyond the equation of value” (2017). Parisi suggests that, instead of being just invisible, blackness, as matter without form, brings forward the nullification of the ocularcentric field of vision.

What constitutes knowledge or value can be seen to be arranged in ways that further recall Berger’s reflections on the medium of television through which his ideas were made public: “But remember that I am controlling and using for my own purposes the means of reproduction needed for these programmes [...] with this programme as with all programmes, you receive images and meanings which are arranged. I hope you will consider what I arrange but please remain skeptical of it.” (1972). We reiterate this statement here to further stress reflection on the means of production, including journals like this, that purport to impart useful knowledge. What is learnt should not be separated from the means by which it is transmitted or circulated. More to the point, the production of meaning lies at the core of our discussion, as do concerns about what is being learnt, and to what extent this has been compromised by the mode of production, or inflected by reductive ideas of how the world operates. Under these conditions, the relations between human and machine learning become quite blurry. The overall idea of “learning” implies new forms of control over what and how something becomes known and how decisions are made, as for instance, in the ways in which images are classified and categorized by humans and machines (e.g. concluding that this image of a person most likely represents a specific gender, race, likely terrorist, and so on). Knowledge is often set at the lowest common denominator in such cases, backed up by the enormous infrastructural power of the companies that profit from generalisations, as is the case for platform-based media empires—such as Amazon and Google—who are concerned that users simply supply (visual) data.

That machines can be said to “see” or “learn” is shorthand for calculative practices that only approximate (or generalise) likely outcomes by using various probabilistic algorithms and models that all have built upon inherent human–machine prejudices such as those related to gender and race (see Myers West et al. 2019). What are the political consequences of automated systems that use labels to classify humans, including by race, gender, emotions, ability, sexuality, and personality? Kate Crawford and Trevor

Paglan’s “Excavating AI” (2019, republished in a new version for this journal) examines how and what computers recognize in an image, and indeed what is misrecognized? Computer vision systems make judgements, and decisions, and as such exercise power to shape the world in their own images, which, in turn, is built upon flattening generalities and embedded social bias.

A cartography of the limits of AI is provided by Matteo Pasquinelli and Vladan Joler in “The Nooscope Manifested” (herein), applying the analogy of optical media to “diagram” machine learning apparatuses. Ultimately, they wish for more collective intelligence about machine intelligence, more public education instead of “learning machines” and their “colonial regime of knowledge extractivism”. The interplay between truth and fiction is again part of this, and “deepfakes” for example (a wordplay on “deep learning”) make a good case study for the ways in which synthetic instances can pass for real data (or human reason) as perpetuated by corporate regimes of knowledge extractivism and epistemic colonialism (to paraphrase the evocative description of Pasquinelli and Joler). Furthermore, this interplay of truth/fake is what Abelardo Gil-Fournier and Jussi Parikka expand upon in “Ground Truth to Fake Geographies” (herein) to chart the development of what they call “ground truth”, with reference to the shift from physical, geographical ground, to the “ground of the image”. They discuss contemporary practices that mobilize geographical earth observation datasets for experimental purposes, including “fake geography” as well as artistic practices, to show how ground truth is “operationalised” (citing Harun Farocki 2004).

Seeing, then, is no longer centred, singular or indexical truth or reality, as it was mistakenly thought to be, indicative of a wider need to manifest authority and power through visibility, but takes on new distributed and contradictory forms. Fabian Offert and Peter Bell in “Perceptual Bias and Technical Metapictures” (herein), argue for a transdisciplinary approach—across computer science and visual studies—to understand the inherent biases of machine vision systems. What they call “perceptual bias” accounts for the differences between the assumed “ways of seeing” of a machine vision system, “our reasonable expectations regarding its way of representing the visual world, and its actual perceptual topology”. This shifts the discussion from critical attention to dataset bias and on fixing datasets by introducing more diverse sets of images. The point is upheld by Nicolas Malev in “On the Dataset’s Ruins” (herein) who acknowledges the dataset as significant cultural form, but also wants to shift attention from fixing the database (as in the case of the inherent biases of ImageNet) to highlight the invisible labour that labels and classifies the images, as well as to the operations of the apparatus itself. To understand the specific character of the “scale” of computer vision as he puts

it, comparison is made to André Malraux's "museum without walls" in which the photographic apparatus allows for unlimited access to the world's image resource. The social implications of this attention to scale—from the molecular to urban—are further elaborated in Benjamin Bratton's commentary on the The New Normal think-tank at Strelka Institute in Moscow ("AI Urbanism: A Design Framework for Governance, Program, and Platform Cognition", herein), thereby understanding AI as "a property that can be designed into objects of different scales" and thus more alert to questions of "what kind of *planetarity* can and must be composed".

From our open call for papers, contributors further address the ways in which existing references from visual culture—such as Berger's *Ways of Seeing*—useful as they remain, require additional work, not least in dialogue with other disciplines to further emphasise the importance of social and political aspects in technical fields such as AI (Agre 1997). (And we might reiterate the name of this journal to stress our point.) Reflecting what he calls "cultural analytics", Lev Manovich's "Computer Vision, Human Senses, and Language of Art" (herein) argues for the importance of using computer vision methods in humanities research, and how this can offer analytical insights into cultural objects where existing analytical tools fall short. In another article, "On Machine Vision and Photographic Imagination" (herein), Daniel Chávez Heras and Tobias Blanke discuss the experimental television programme *Made by Machine: When AI met the Archive* (2018), created with materials from the BBC television archive using different computational techniques. They establish a conceptual and material link between photographic practice and deep learning computer vision, and the implied "optical perspective of the computer vision system itself". Carloalberto Treccani's "The Brain, the Artificial Neural Network and the Snake" (herein) is an account of the evolutions of vision systems, from "intelligent" animals to the functioning of Artificial Neural Networks. Both human and machine learning can be seen to demonstrate, through trial and error (recurrent neural nets), how, for better or worse, artificial systems inform us about the workings of biological systems, and how and "why we see what we see". Claudio Celis Bueno and María Jesús Schultz Abarca's article "Memo Akten's Learning to See" (herein) employs the philosophy of Bernard Stiegler to insist that "human vision is always already technical", and the result of constant training processes. Akten's art installation *Learning to See* (2017) becomes an example of machinic imagination and "machinic unconscious" (referring to Walter Benjamin's notion of the "optical unconscious", and how technical reproduction provided access to new forms of vision). The optical logic is transformed, in ways that augment new layered realities. This layering is what Manuel van der Veen describes, in his "Crossroads of

Seeing" (herein), in which two ways of seeing are produced simultaneously, as our field of vision is superimposed with additional information and images. Augmented reality is compared to traditional procedures, such as *trompe-l'œil*, to suggest that new kinds of reflection become possible; not only to look at the intersection itself, but also to see where the ways "divide" (we might add contradict).

The inherent fallibility of AI systems is explored in Gabriel Pereira and Bruno Moreschi's "Artificial Intelligence and Institutional Critique 2.0" (herein), drawing upon their intervention within the art collection of the Van Abbemuseum in Eindhoven, Netherlands. Using widely available image-recognition software to "read" images, the inherent values of both are exposed. It is the "untrained eye" of computer vision that offers critique of the art system and possible new ways of approaching visual culture on the one hand, as well as understanding the commercial imperatives of computational ways of seeing on the other. The point is reiterated that new methods of analysis are required that work with computational practices and existing theoretical readings, in which the human and the machine are critical partners. In Iain Emsley's "Causality, Poetics, and Grammatology" (herein), a "critical assemblage" of philosophy and computational thinking allows for an analysis of the *Next Rembrandt* project (2016, a 3D printed painting made from the data of Rembrandt's total body of work using deep learning algorithms and facial recognition techniques). In Rebecca Uliasz's "Seeing like an Algorithm" (herein), the relationship between the image and the human subject is explored in detail. Techniques of machine vision are described as "techniques of algorithmic governance", and the way the human subject is made visible through computation. She refers to the relationship between the "operative image" (Farocki 2004) and the formation of "emergent subjects". Perle Møhl's "Seeing Threats, Sensing Flesh: Human-machine ensembles at work" (herein) further develops the discussion of "human-machine ensembles", how they work together mutually to "see" specific things and "unsee" others. Her key point is that seeing in both cases is "not automated but unskilled and mutually co-constituted". Examples are provided that demonstrate this coming together of "material, political, organisational, economic and fleshy entities in order to configure what can be seen and sensed, and what cannot".

We hope with this coming together of articles there is sufficient attention to critical-technical practices that illuminate the complexity of human-machine relations, and their transformations, and not least to serve to emphasise the uncertainties and contingencies that characterise these contemporary ways of seeing.

The relation between what we see and what we know has always been unsettled. It has been argued that seeing and ways of seeing have been intertwined since bipedism

bifurcated the evolution of the first anthropoids, which found themselves with hands for crafting (rather than only using them for walking) and a mouth for speaking (rather than only for eating) (Leroi-Gourhan 1993). The changing of orientation happens simultaneous to the development of an erect posture and the consequent changed function of the mouth, from an organ to simply gather and ingest food, to an organ increasingly capable of articulating sounds. Building on the philosophy of André Leroi-Gourhan, it is possible to imagine that this changing of perspective stimulated a vocal response from the liberated mouth—and the sequence of sounds were a first attempt to articulate the sense of awe (Plato 2004) or dread (Nietzsche 2006) associated with the new orientation of the body. At the same time, the liberated hands could start intervening with the surrounding landscape, scratching stones and, in a way, attempting inscription of those early articulations of sound emerging from the liberated mouth. Otherwise said, the technical ability acquired by the hands proceeds simultaneous to the possibility of articulating sounds to signify things, thanks to the new orientation of the erect body, and the freeing of the head and the mouth.

Although Berger seems to be strongly influenced by phenomenology—especially the phenomenology of perception of Merleau-Ponty (1964), and defines human seeing as primordial and in anticipation of language—the relation between seeing and ways of seeing is unsettled because seeing (as much as speaking) is always already technical, and, as such, written (Derrida 1976). Despite these phenomenological traces found in Berger’s approach, his work is also about the hidden intricacies between image and language, explored through a reverse engineering of the grammar of images (no doubt influenced by the influence of semiotics and structuralism at that time). Berger maintains the primordality of the image over the language, as is made clear in the first sentence of the book—“Seeing comes before words. The child looks and recognizes before it can speak”—or in the assertion that although we “explain that world with words [...], words can never undo the fact that we are surrounded by it” (Berger 1972: 7). Furthermore, “the reciprocal nature of vision [of simultaneously seeing and being seen] is more fundamental than that of spoken words” (Berger 1972: 8). Yet, to Berger, it is the very grammatical in-expressibility of the image that inevitably generates the power of signification through sounds and language, which in our earlier brief sketch of evolutionary palaeontology (inspired by Leroi-Gourhan) began with the changed orientation of the body, producing new ways of seeing together with inscription practices.

Moreover, Berger turns the phenomenological ideal of a primordial image into a trigger for a historical materialist analysis of images which departs from the idealism of a virgin gaze and white canvas, and instead explores the originarity of the image in relation to the proliferation of

strategies for registering and manipulating it. It’s not difficult to imagine these early anthropoids emitting sounds while simultaneously experimenting with modifying rocks into what became tools. In this sense, Berger’s phenomenological and semiotic references, find their current synthesis in a post-phenomenological approach which rethinks the originarity of the (human) image as always already captured by the originarity of technicity (Stiegler 1998), which allows its perception in the first place, turning the image into a way of seeing.

In this sense, “all images are [hu]man-made” (Berger 1972: 8), and all images imply knowing in their making as much as in their realisation—this knowing, and its temporality, being inscribed, or spatialised, in the proto-technical object, if we follow Stiegler. Thus, not only “the way we see things is affected by what we know or what we believe” (Berger 1972: 8), but also the images we make (or how we make them) are “affected by what we know or what we believe”. It seems there is always a grammar beyond an image, and this grammar proliferates because of its impossibility to grasp the fullness of an image. The relation between image and language is necessarily unsettled because, although they are co-originary, the ability of the latter to signify the former is never settled, and changes as a consequence of what we know or believe.

Since the beginning, the unsettling relation between (human-made) image (ways of seeing) and language (knowing) has turned into the unsettled relation between *techne* (the *techne* to make an image) and *episteme* (the socio-cultural milieu that allows for certain *technics*, and images, to emerge). In fact, the relation between *techne* and *episteme*—or, in Berger’s terms, seeing and knowing—is so deeply unsettled that philosophy, according to Derrida (1981) and Stiegler (2013), has been unable to think it properly since at least the time of Plato’s *Phaedrus* (370 BC). At this time, the unsettling relation between ways of seeing as always technical and knowing, and as always supported by technicity, is exemplified by the practice of “sophism”, understood as a break of the linkage between *techne* and *episteme*. In brief, sophists applied *techne* without a proper *episteme*, turning it into a poison, and not a cure, given the pharmacological nature of technology, already highlighted by Plato, and further emphasised by Derrida and Stiegler. Berger’s work also dives into this de-linkage, not at the level of philosophical inquiry, but at the level of the concrete production of artefactual images and the implicit process of knowing they support (or not). In doing so, Berger provides tactics for navigating their hidden epistemologies, within a Marxist framework in the most literal sense, through which to re-think the relation between the technical exteriorisations which produce images and the concept of alienation as the distancing-effect between the means of production and the knowledge of conditions. Alienation here is understood as

a departure from the relation between the proliferation of technical exteriorisations and the implicit (gnoseological and) epistemological framework they bring forth.

These inscriptions, or technical artefacts, are written, and represent images. Technical exteriorisations (images, in Berger's sense) attempt the inscription of the non-specialised sensory-motor schema of the hands and of the erect body which, as a consequence of its non-specialisation, produces the corticalisation of the brain (Leroi-Gourhan 1993). Non-specialization through externalisation opens up the therapeutic or curative side of the technical object, while its poisonous side emerges when externalisation starts to produce, instead, specialization. Factory workers are specialised because they only know how to operate the cog they see in front of them—without knowledge of the larger mechanism of which it is part—as dramatically shown in Farocki's (1968) film *Inextinguishable Fire*, depicting the production of Napalm by Dow Chemicals by units of workers producing single parts necessary to assemble the weapon. Workers remain unaware of the ultimate goal of their work, given the strict division of labour enforced by the factory. The distancing between seeing and knowing increases with computational technology, which often reduces cognitive workers into keyboard operators, pushing buttons completely unaware of the algorithmic consequence of their actions. The invisibility of the grammar is triggered by computational infrastructures that turn ideology (and knowledge) into a cloud of diffuse, networked services constantly extracting and processing data. This invisibility is a consequence of the specialization of knowing and it is de-linking from the ways of seeing that enable it in the first place. This alienation, or “generalized proletarianization” (Stiegler 2017) involves producers and consumers alike, and their common loss of *savoir-faire* (how-to-do), *savoir-vivre* (how-to-live) and *savoir-théorizer* (theoretical knowledge). Rethinking the linkage between knowing and seeing in relation to ways of machine seeing means understanding alienation as the theft of knowledge operated by technics animated by sick epistemological frameworks, and vice-versa. Furthermore, it means to understand that the technical object has been deprived of its ability to produce new long circuits of individuation and trans-individuation (Stiegler 1998), and instead, as we will see, produces short circuits of dividuation.

These tendencies proliferate and become more complex as the relation between visibility and invisibility keeps shifting and reaches a point where the demand to visualise drives a bulimic “drive to visibility” which aims at making everything visible (van Winkel 2005: 1). At the same time, the drive to visibility functions by hiding the processes through which visibility emerges. In the current technological milieu, big data is funneled from raw data assemblages into datasets that furnish the materials for the constitution of users'

algorithmic doubles (which arise from the extraction of their geolocation, frequency of communication, choice of topics, and so on). In this way, the algorithmic double becomes a data matrix for the molecular-tailored production of “missing visuals” (van Winkel 2005). In this architecture, missing visuals are those that appear at the level of the interface on the basis of the user's algorithmic double, and function as bait to keep the user clicking and producing new data to enrich the algorithmic double which, in turn, will produce new (missing) visuals. In short, the proliferation of big data's invisibilities function as means for the production of new data and, as a consequence, new missing visuals (Azar 2020). These invisibilities capture the subject in a circuit of algorithmic dividuations which produce the subject's algorithmic double, designed to overlook the production of (missing) images, which is where the circuit of individuation (and trans-individuation), possibly enabled by the technical object and its ways of seeing, is interrupted. The projection of the algorithmic double on the user's interface generates mirroring effects and digital echo-chambers where previous tastes and beliefs are fed back to confirm their positioning (and orientation) in the world. In Berger's terms, and put simply, new ways of algorithmic seeing allow the constant production of new visuals while simultaneously increasing the gap between seeing and knowing.

If this relation between seeing and knowing was once fundamental to acting in the world, the current distribution of agency across complex networks of non-human agents allows simultaneously more visibility—and, as a consequence, more knowledge about processes that before were not visible—and less knowledge about the very processes behind the way in which these new visualities are rendered visible. Although we see more, we are not given the instruments to understand the ways in which we see more—this being mainly a problem of property, for example in relation to our reliance on proprietary platforms and infrastructures, but also a problem related to the complexity of algorithmic networks, which in their operations often escape current epistemological frameworks.

Paraphrasing Stiegler (2015, 2019), if we see (and know) more, it is because of the computational ability of algorithmic networks to take over (and exponentially increase beyond human capacity for reason; what Kant defines as the analytical faculty of reason, or understanding). If we see and know less, simultaneously, it is because algorithmic networks colonise also the faculty of synthesis, “short-circuiting the deliberative functions of the mind” (Stiegler 2019: 26), to the point of moving beyond the given epistemological framework which embed them, or, otherwise said, to the point of destroying the possibility of theory (and of a human-accessible epistemology) as such (Stiegler 2015). In this context, how to recover a sense of agency when it has

been distributed to wider systems and assemblages, or more to the point, how to adjust to this reality (Tsing 2015)?

Evidently, images themselves have the ability to act in the world, and upon us, nowadays via the artefactuality of computational technology, characterised by its simultaneous invisibility and proximity. According to Stiegler, adjusting to reality consists of adopting a given (artefactual) reality rather than adapting to it (2010), with adoption standing for the ability to recover a sense of agency through long circuits of individuation that increases knowledge as *savoir-faire* (2018) and non-specialisation via the adoption of the technical object and its emancipatory potential. In a way, this was once the project of critical visual culture, to which Berger's essay contributes: to render visible the underlying conditions that allow us to see reality as it really is, instead of bringing to visibility new visuals by hiding the processes that make them visible, as happens with AI-driven big data analysis for example. Moreover, this is about the relation between what is visible and the names that we give to what is seen, as well as what is invisible—a politics of representation and nonrepresentation. But things are not so easy to comprehend as they were once thought to be, as images now proliferate and circulate in such vast quantities and are mostly made by machine for machines (Paglen 2016), and the knowledge related to this circulation can be only partially traced given the ability of machines to write according to a grammar not fully graspable by humans, further occluded by the proprietary nature of this form of writing, and knowing.

That machines play their part in this multilayered assemblage of images, turning images into operations, was famously articulated by Farocki: “images no longer represent an object but are part of an operation” (2004). In his video trilogy *Eye/Machine* (2001–3), and in his writing “Phantom Images”—key references for various articles in this journal—this is made evident, as the image-making machine no longer simply takes the position of a person but one of “intelligence”, combining what Farocki calls “the ill-considered notion of intelligence with an equally ill-considered subjectivity” (2004: 13) Images then are “operative” inasmuch as they do more than simply display themselves and offer themselves for human interpretation, but begin to interpret themselves. Machines “see”, but not simply like the eye of modernism (as in the case of the “kino eye” of Vertov's 1929 *Man With a Movie Camera*). Image-machines “act”, but no longer in the metaphorical or animistic “image-act” that art historian Horst Bredekamp saw in the war-photograph or the medieval altarpiece. In their social interaction with humans, images have always created, and not just depicted, reality. Now, as in the case of images created by machines for machines (synthetic datasets, QR codes, calibration

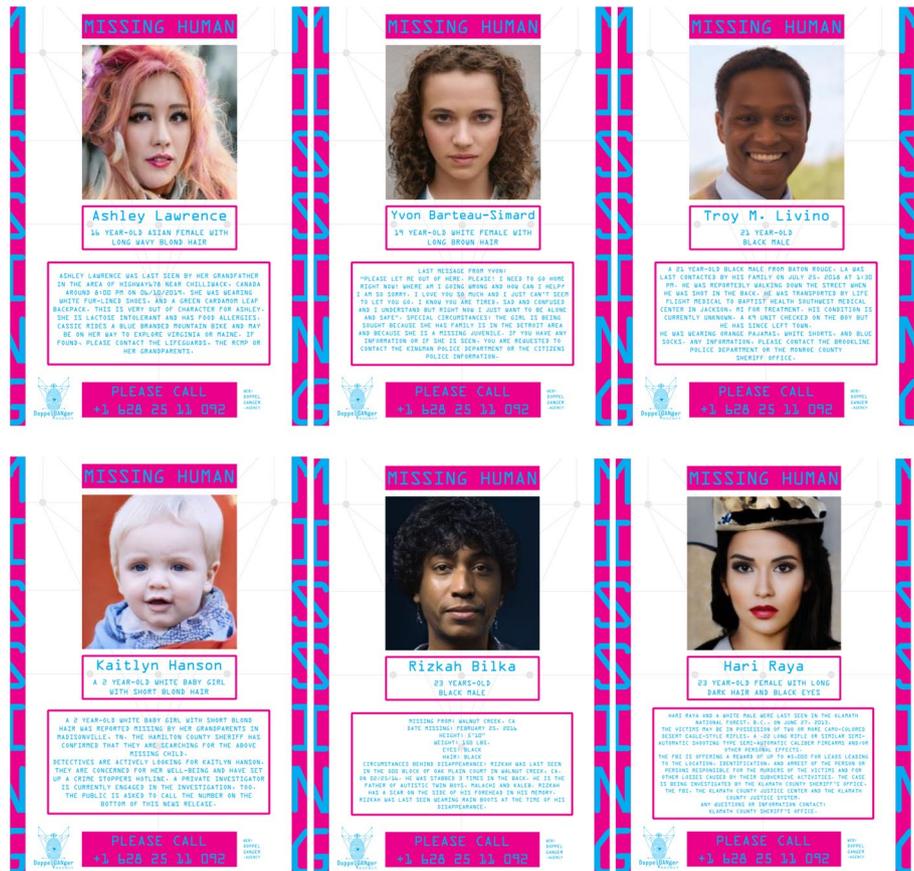
routines, and so on), image-machines create reality in their autonomous interaction with each other.

The removal of the human subject from the assemblage of seeing, or its role as a disposable element from which to extract value and knowledge, is symbolised by what Farocki calls the “suicidal camera” (2004). Suicidal cameras generate suicidal images: images produced by cameras located on remote missile systems and offered to the human operator until the moment the missile hits its target, at which point the camera disintegrates with the explosion and the transmission ends. If on the one hand, the human element fulfills a monitoring function, while the (ill) intelligence is delegated to the machine, on the other it is the pray of the image, and appears and disappears with it. Farocki's suicide images work well to explain the relation between the iconic indexicality of the algorithmic image understood as the possibility of its human-oriented use-value (otherwise said, as the form in which it is accessible by humans), and the speed at which it is forced to circulate and to be exchanged (according to its exchange value), which, in fact, compromises its human-oriented (iconic) indexicality.

It is within this type of operability—understood by expanding on Berger's Marxist framework—that the drive to visibility bargains truthfulness for novelty, and the relation between seeing and knowing becomes a problem of truth, or, to say it with Foucault, a problem of “games of truth” (Lorenzini 2015). In our computational context, iconic indexicality falls short not really because algorithmically-produced images look fake, rather the opposite—they can almost seamlessly pretend to be factual, indexing reality while being fully algorithmically-generated. For example, Generative Adversarial Networks (or GANs)—a framework in which two neural nets are dialectically opposed—have managed not only to allow real time facial re-enactment (as in the case of the DeepFakes) but also to process autonomously huge databases of real human faces and to generate new hyper-realistic faces that do not replicate any of the faces of the dataset. These AI human faces are both faces of missing humans (who do not exist in the actual world) and faces of algorithmically-generated ghosts, as paradoxically shown by the project *DoppelGANger.agency* (Azar 2018), which turns AI-generated faces into street posters for missing persons (Fig. 2).

These DoppelGANgers are a form of missing visuals emerging from the real human faces the GANs were trained on (Azar 2020), and as such can help to re-open the question of iconic indexicality in relation to algorithmic images that resemble their referents, even though their referents do not exist. In other words, the iconic indexicality of the algorithmic image serves to support the algorithmic indexicality which is at the core of the operational drive and consequent circulation of the image. This algorithmic indexicality allows for the clustering of images with similar parameters,

Fig. 2 Examples from Mitra Azar, Doppelganger.agency (2019), doppelganger.agency



and to profile users exposed to the iconic indexicality of the image on the basis of those clusters, so as to predict the next (missing) image. In fact, the seeing and knowing on the side of the algorithmic indexicality is almost alien to the seeing and knowing happening on the side of the iconic indexicality, yet, at the same time one would not function without the other.

The relation between ways of seeing and knowing is always unsettled then because although the former is a grammatical object which, on the basis of its grammatology (of which the technical inscription is a trace), manifests itself as a knowable object, its very grammatology doesn't exclude—but rather depends on—the incomputable opening that the relation between ways of seeing and knowing inevitably produce. Ways of seeing and knowing—technical artefacts and human knowledge—are rooted in the non-specialised senso-motor and cortical potential defined by its capability of bifurcating, or finding ways of *savoir-faire* and *savoir-vivre* not fully inscribed by the ways of seeing and the forms of knowing they enable.

Finally, in this last section, we should say more about our editorial process and motivation for our work. We started this project with a series of small events at the University of Cambridge in 2016 (initially developed by Anne Alexander, Alan Blackwell, Geoff Cox and Leonardo Impett), to bring

some of the politics of Berger to the discussion of machine vision, working across the Digital Humanities Learning programme and Computer Lab, and reaching out to other collaborators, and in turn to the contributors of this journal. Our starting point was to speculate on the comparison between Berger's *Ways of Seeing* and David Marr's *Vision*, a foundational text in computational neuroscience—and the axiomatic philosophy of vision behind machine vision for the past four decades. If Berger's concern was understanding how humans seeing with machines changed the ways in which they could represent the world, Marr was interested in the theoretical work necessary to make machines which could see. With a degree in mathematics, his early work on neuroscience was largely done at Cambridge, before moving to MIT; where in 1980, he was tenured at the Department of Psychology. He died later that year. His book, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, was published posthumously two years later, becoming the foundational text for the new field of computational neuroscience.

The explanation of the human visual system contained in Marr's *Vision* has remained the go-to theory of human vision for generations of computer scientists, for teaching and research: the most prestigious award in the computer vision community (awarded by the International Conference

on Computer Vision) is called the *Marr Prize*. What must be stressed, however, is that computer scientists didn't just adopt Marr's paradigm because it was the most up-to-date, successful, or accurate physiological theory of the human visual system. Computer vision is not, in general, meant to be a simulation of a biological visual system; that is not the epistemic role of a theory of human vision for the computer vision community. Rather than giving a biological precedent that machine vision researchers could copy, Marr's theory does almost the reverse: it understands human visual processes *in computational terms*. The parallels Marr draws between human and machine vision are far wider, and more structural, than that of the individual "neuron". Rather than anthropomorphic computing, Marr's *Vision* is technomorphic physiology.

Marr was interested in creating an information-theoretical (and even computer-scientific) model of biological seeing: before *Vision*, his early work was to propose computational models of the cerebellum, neocortex and archicortex (hippocampus) (Marr 1969, 1970, 1971). These were computational models not in the sense of computer simulations, but rather in that they used the metaphors of theoretical computer science—pattern recognition, memory systems—to explain the role and mechanisms of the human visual system. Electronics had long furnished a conceptual and notational system with which to describe human neuroscience (as in the Fig. 3 below, from Marr 1982); Marr's innovation was to extend this metaphor to information theory itself (Fig. 3). This is the basis of Marr's central axiom: *vision as an information processing system*.

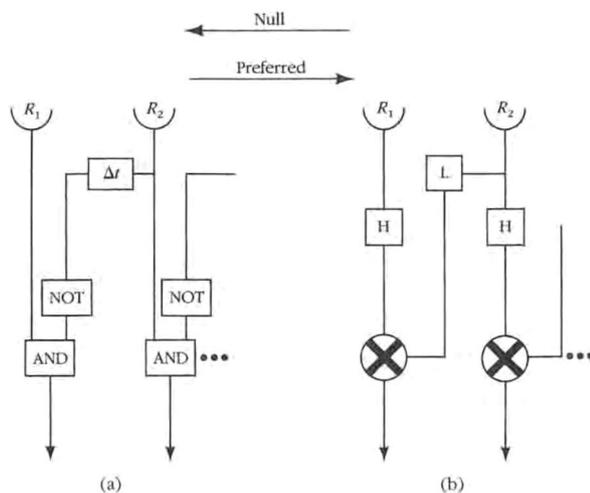


Figure 3–30. (a) Barlow and Levick's (1965) model for directional selectivity connects two detectors to an AND–NOT gate, one via a delay. Thus the network does not respond to stimuli moving with roughly the right speed in the null direction. (b) Hassenstein and Reichardt's (1956) model operates on the same principle except that the delay is replaced by a temporal low-pass filter (L). H = high-pass filter.

Fig. 3 Different models (in Marr, *Vision*, 1962: 163)

In the decades that followed Marr's *Vision*, a number of results supported his general theory of vision as information processing: that the human trichromatic system (Parkkinen and Jaaskelainen 1987), and the activation functions of primary visual cortex neurons (Field 1987), could be derived statistically from data about the natural world. In the so-called *Tri-Level Hypothesis*, David Marr and Tommaso Poggio (1976) outlined how these information processing systems—both biological and machine vision—ought to be understood at three separate levels:

- Computational level: what (computational) problem does the system solve, and why?
- Algorithmic/representational level: what statistical representations and operations does the system employ to solve this problem?
- Implementation/physical level: how are these statistical operations implemented, either as neural structures or computer code?

Marr's proposal, therefore, is to understand machine and human vision as a separable *stack*: we might share the same computational problem (say, face recognition) across multiple algorithms (eigenfaces, wavelet transforms), or the same algorithmic approach (e.g. learning sparse feature embeddings), through various physical implementations (neurons, GPUs, CPUs). Though fond of computational metaphors, Marr might not have recognised the concept of the *software stack*, which only really took off with the introduction of the OSI model for telecommunications in 1984 and the 2000s proliferation of stack models for web technology. Software systems are increasingly designed, conceived, operated, and programmed around the metaphor of the stack: not just from a technical perspective ("full-stack developer"), but also in their political-economic dimension ("full-stack business analyst").

Bratton's intuition (2016) about the "stack"—that the structural, transversal metaphors of software engineering and techno-capitalism are equally useful critical tools—is relevant to our project. We can expand Marr and Poggio's tri-level model of human vision:

- (1) Social level (where are such systems deployed, by whom, for what purpose)
- (2) Computational level (which problems are being solved: e.g. "object detection")
- (3) Data level (who labels, which images are chosen, who takes the photographs)
- (4) Algorithmic/representational level (e.g. Siamese convolutional neural network with Adam gradient descent optimization)
- (5) Implementation/physical level (abor. Tensorflow on cuDNN/CUDA on Nvidia GPU)

(6) Philosophical/axiomatic level (e.g. *vision as inverse graphics*)

This vertical cartography of machine vision—though far coarser than Joler and Pasquinelli’s Nooscope—nonetheless exposes black-spots in our collective critical dissection of the machine vision stack. We have brought forward an excellent critical understanding of the datasets of machine vision and of their social applications, but have almost nothing to say about the ideological content of specific algorithms, technical axioms, compiler languages, or massively parallel silicon implementations.

One of these technical axioms comes from the work of Marr himself: the computational paradigm of [robotic] *Vision as Inverse* [computer] *Graphics*. Marr outlined what he saw as the three stages of vision in an information-processing pipeline in the construction of:

- (1) Firstly, a 2D primal sketches: including edge detection, silhouettes, etc.;
- (2) Subsequently, 2.5D images, including textures, foreground and background;
- (3) Finally, a full 3D model of the environment.

Here, the machinery of machine vision is chained to that of the early video-games industry (a link which extended, since widespread use of CNNs in 2012, to the hardware level of the stack—how Nvidia overtook Intel); through the implicit separation of mesh and texture, object and background, image and sprite. This is what makes it possible to create *synthetic datasets* in machine vision; to use CGI to synthesise models of faces, cars, streets, or warzones, which machine vision algorithms then learn from.

Where Berger sets out a dialectical relation between seeing and knowing, and the role of knowing in seeing, Marr’s suggestion is that the visual system is fine-tuned to efficiently compress (i.e. recognise) visual stimuli that it has evolved to encounter (faces, shadows, motion); and in doing so has a kind of implicit, embodied knowledge about the natural world and its images. For Marr, then, and for computer vision scientists after him, seeing is not so much a product of *knowing* (nor, we might add, of *belief*, and thus ideology) as a product of *data*. A consequence of this perspective—of the unmediated nature of visual representation—is an inability to deal with the ambiguities and inconsistencies of visual perception, whose technical implications have been highlighted by Aaron Sloman, former chair of AI at Birmingham University: “that common idea is mistaken: visual systems do not represent information about 3-D structure in a 3-D model... but in a collection of information fragments, all giving partial information. A model cannot be inconsistent. A collection of partial descriptions can” (2011). Sloman’s

critique of Marr here echoes Berger: “the relation between what we see and what we know is never settled.”

As Berger wrote of photography, the technical aspects of machine vision “are not, as is often assumed, a mechanical record”, rather they are saturated with ideology. What is to be done? Taking political issue with the corporate machinery of machine vision is, for at least two reasons, an intrinsically slippery task. First, because this corporate machinery is difficult to delineate; it’s as far as it’s possible to get from the perfect-substitute-producing factory of textbook economics (Srnicsek 2016). It includes conventional private companies which sell machine vision for profit—but also corporate-funded AI labs which publish research openly in the scientific community; privately-funded open-source software libraries, on which much of AI depends; and infrastructure, from cloud computing to hard silicon. Google does all of these; selling machine-vision-as-a-service (Google Vision AI), publishing open research (Google Brain), producing fundamental shared open-source libraries (Google TensorFlow), and selling the cloud services (Google Cloud) and even the chips (Google’s Tensor Processing Units). Microsoft, Amazon and others have similar profiles.

But there is a second reason for which the intersection of political critique and corporate research in machine vision is so confusingly nuanced. The academia-versus-industry debates of the past five decades had clear demarcation lines; for instance, in the tobacco industry’s long refutation of the link between smoking and cancer, or the oil industry funding research which attempted to obscure the links between man-made emissions and climate change. In the case of machine learning and machine vision, industry (including corporate AI labs) publicly agrees, in many cases, with its critics. Microsoft, for instance, claims “inclusiveness” and “fairness” as two of the six guiding principles for Responsible AI. This is not simply a marketing proposition—a paper on fairness in AI from Microsoft Research won Best Paper at CHI 2020 (Madaio et al 2020), and similar internationally-relevant research is produced by other major corporate tech players. Through efforts like increasing demographic diversity in the workplace, technology companies have endlessly publicised those occasional win-win situations in which a degree of corporate fairness increases long-term profitability. But it’s not yet clear to what degree this mandate for socially—and ethically responsible machine learning extends to activities which might seriously endanger a company’s bottom line: well-publicised ethical AI policies notwithstanding, Amazon, Google, Microsoft, Oracle and IBM all still bid for a \$10bn cloud computing contract with the US Department of Defense in 2018–19. When Google dropped its bid in October 2018, it claimed the first reason behind this decision was that “we couldn’t be assured that it would align with our AI Principles”; in practice, it was responding to seven months of significant organised pressure

from Google workers, including several resignations and a widely-signed employee petition.

Where policymakers and journalists tend to talk about analysis of abstract *data* (almost an implication of Excel spreadsheets), a very large proportion of today's AI controversies centre around *vision*. The debate over Google's involvement in the DoD contract started when it was revealed that Google supplied machine vision technologies for the automated video analysis of drone footage. Amazon (who, at time of writing, was attempting a legal challenge to the contract being awarded to Microsoft) continues to publicly host an online demo of Amazon Rekognition, their machine vision platform, for the defence industry; highlighting Amazon's ability to "analyze images and recognize faces, objects, and scenes", and finding the "likelihood that faces in two images are of the same person, in near real-time". Without a hint of irony, Amazon is at the same time the principal sponsor of a 3-year \$20 million funding program of the U.S. National Science foundation on Fairness in AI.

Such paradoxical situations, in many cases, are not borne of inconsistency or hypocrisy, but *irresolvability*—a general symptom, for Matthew Fuller and Olga Goriunova (2019), of the post-Cold War society. In a structurally unequal society, it is exceedingly difficult to make a "fair" algorithm; and it is effectively impossible to make an algorithm which is both fair *and* effective. Jon Kleinberg, Sendhil Mullainathan, Manish Raghavan (2016) have shown, in the most general case it is *impossible* to classify data in a way that meets several well-established definitions of fairness (lack of active discrimination; balanced false positive rates; balanced false negative rates). All three definitions of fairness also produce classifications which are, in general, different from the highest possible accuracy (and therefore, in most situations, different from the highest potential generated profit). There are two special cases where it is possible to satisfy all fairness conditions: *perfect predictions* (total information—nobody is ever misclassified because of their background), or *equal base rates* (i.e. a society/ground-truth where different groups have statistically identical behaviours). In a society which is unfair, a classification-machine will *always* be unfair (in at least one sense). What is so important about Kleinberg et al.'s mathematical proof is that it's fundamentally about the final *classification results*, regardless of how any individual decisions are taken. The impossibility of fair classifiers in an unfair world, therefore, is equally true if those classifiers are human. In the specific case of machine vision, the risk of unintended discrimination through high-dimensional correlation has greatly increased with the advent of deep convolutional neural networks (Impett 2018), compared to their simpler (and less powerful) predecessors which relied on hand-crafted geometric features.

Nonetheless, the research behind such a profound mathematical result—with enormous political implications for human decision-making—would not have existed were it not for current debates around machine-learning-based classification systems. Our hope, in this journal and beyond, is that critical work on machine vision will lead to similarly profound political insights. "To ask whether machines can see or not is the wrong question [...] rather we should discuss how machines have changed the nature of seeing and hence our knowledge of the world" (Cox 2016). In this sense, the project of algorithmic literacy behind *Ways of Machine Seeing* mirrors Berger's didactic project of visual culture.

Beyond being a powerful or harmful new technology in its own right, machine vision gives us a new, precise set of metaphors with which to think about vision differently—in computational, biological, aesthetic, and consequently *political* terms—to enhance our ability to see and thereby act in the world.

Acknowledgements *Ways of Machine Seeing* first emerged as a workshop organised by the Cambridge Digital Humanities Network, convened by Anne Alexander, Alan Blackwell, Geoff Cox and Leonardo Impett, and held at Darwin College, University of Cambridge, on the 11 July 2016, and which has involved many more people and institutions since. We wish to thank our reviewers—without whom the journal would not have been possible—for their work, their detail and their rigour: Alan Blackwell, Annet Dekker, Andrew Dewdney, Jonathan Impett, Paul Melton, Gabriel Menotti, Darío Negueruela del Castillo, Winnie Soon, and Pablo Rodrigo Velasco González.

Most of all, we must thank the authors contained in this journal, for the many hours of uncompensated labour, and for the quality of their articles: María Jesús Schultz Abarca, Peter Bell, Tobias Blanke, Benjamin Bratton, Claudio Celis Bueno, Kate Crawford, Iain Emsley, Abelardo Gil-Fournier, Daniel Chávez Heras, Vladan Joler, Nicolas Malevé, Lev Manovich, Nicholas Mirzoeff, Perle Møhl, Bruno Moreschi, Fabian Offert, Trevor Paglan, Jussi Parikka, Luciana Parisi, Matteo Pasquinelli, Gabriel Pereira, Carloalberto Treccani, Rebecca Uliasz, and Manuel van der Veen.

References

- Agre PE (1997) Toward a critical technical practice: lessons learned in trying to reform AI. In: Bowker G, Gasser L, Star L, Turner B (eds) Bridging the Great Divide. Social Science, Technical Systems, and Cooperative Work, Erlbaum
- Azar M (2019) Pov-data-doubles, the dividual, and the drive to visibility. In: Natasha L (ed) Big data—a new medium? Routledge, London, pp 177–190
- Azar M (2018) DoppelGANger.agency. Available from: <http://doppelganger.agency/>.
- AWS (2017) Amazon Rekognition Demo for Defense. Blog, August 7. <https://aws.amazon.com/blogs/publicsector/amazon-rekognition-demo-for-defense/>
- Benjamin W (2008) The work of art in the age of mechanical reproduction. Belknap Press of Harvard University Press, Cambridge
- Berger J (1972) Ways of Seeing. BBC and Penguin.

- Crawford K, and Paglen T (2019) Excavating AI: The Politics of Images in Machine Learning Training Sets. www.excavating.ai
- Cox G (2016) Ways of Machine Seeing. *Unthinking Photography*. unthinking.photography/articles/ways-of-machine-seeing
- Farocki H (2004) Phantom Images. *Public* 29:12–22
- Farocki H (1968) *Inextinguishable Fire*. Film.
- Derrida J (1976) *Of Grammatology*. John Hopkins University Press, Baltimore
- Derrida J (1981) *Disseminations*. Athlone Press, New York
- Ferreira da Silva D (2017) “1 (life) ÷ 0 (blackness) = ∞ – ∞ or ∞ / ∞: On Matter Beyond the Equation of Value.” *e-flux* 79, February. www.e-flux.com/journal/79/94686/1-life-0-blackness-or-on-matter-beyond-the-equation-of-value
- Field D (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am* 4(12):2379–2394
- Fuller M, Goriunova O (2019) *Bleak Joys: Aesthetics of Ecology and Impossibility*. University of Minnesota Press, Minneapolis
- Leroi-Gourhan A (1993) *Gesture and Speech*. MIT Press, Cambridge
- Lorenzini D (2015) “What is a regime of Truth?”. *Le Foucauldien* 1/1. Open Access Journal for Research along Foucauldian Lines. 2015.
- Impett L (2018) Artificial intelligence and deep learning: Technical and political challenges. *Theory & Struggle* 119:82–92
- Kleinberg J, Mullainathan S, and Raghavan M (2016) Inherent trade-offs in the fair determination of risk scores, arXiv preprint
- Mackenzie A, Munster A (2019) Platform seeing: image ensembles and their invisibilities. *Theory Cult Soc* 26(5):3–22
- Madaio M, Stark L, Wortman Vaughan J, and Wallach H (2020) Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI, 2020 CHI Conference on Human Factors in Computing Systems.
- Made by Machine: When AI met the Archive (2018) Dir. Hannah Fry. BBC Four, September 5th, <https://www.bbc.co.uk/programmes/b0bhwk3p>.
- Marr D (1982) *Vision: a computational investigation into the human representation and processing of visual information*. MIT Press, Cambridge
- Marr D, Poggio T (1976) From understanding computation to understanding neural circuitry. MIT Technical Report, Cambridge
- Mbembé A (2003) *Necropolitics*. *Public Cult* 15(1):11–40
- Merleau-Ponty M (1964) *Phenomenology of perception*. Routledge, New York
- Mirzoeff N (2011) The right to look. *Crit Inq* 37(3):473–496
- Myers West S, Whittaker M, and Crawford K (2019) *Discriminating Systems: Gender, Race and Power in AI*, AI Now Institute, New York University, ainowinstitute.org/discriminatingystems.html
- The Next Rembrandt (2016) Wunderman-Thompson/Microsoft, www.nextrembrandt.com
- Nietzsche F (2006) *On the genealogy of morality*. Cambridge University Press, Cambridge
- Paglen T (2014) “Operational Images.” *e-flux journal* #59, www.e-flux.com/journal/59/61130/operational-images
- Paglen T (2016) Invisible images (your images are looking at you). *The New Inquiry*. <https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/>.
- Parkkinen J, Jaaskelainen T (1987) Color representation using statistical pattern recognition. *Appl Opt* 26(19):4240–4245
- Plato (2004) *Meno*. Focus Publishing, Newburyport, MA
- Sloman A (2011) “What’s vision for, and how does it work? From Marr (and earlier) to Gibson and Beyond.” Birmingham Vision Club, www.cs.bham.ac.uk/research/projects/cogaff/talks/sloman-beyond-gibson.pdf
- Stiegler B (1998) *Technics and Time I-II-III*. Stanford University Press, Stanford
- Stiegler B (2010) *Taking care of youth and the generations*. Stanford University Press, Stanford
- Stiegler B (2013) *What Makes Life Worth Living: On Pharmacology*. Polity Press, Cambridge
- Stiegler B (2015) *States of shock: stupidity and knowledge in the 21st century*. Polity Press, Cambridge
- Stiegler B (2017) *Automatic Society*, vol I. Polity Press, Cambridge
- Stiegler B (2019) For a neganthropology of automatic society. In: Pringle T, Koch G (eds) *Machine*. Meson press, Lüneburg
- Srnicek N (2016) *Platform Capitalism*. Polity, London
- Tsing AL (2015) *The Mushroom at the End of the World: On the Possibility of Life in Capitalist Ruins*. Princeton University Press, Princeton, NJ
- Van Winkel C (2005) *The Regime of Visibility*. NAI, Rotterdam

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.