

RESEARCH ARTICLE

Military techno-vision: Technologies between visual ambiguity and the desire for security facts

Rune Saugmann* 

University of Tampere, Finland

*Corresponding author. Email: rune.saugmann@uta.fi

(Received 10 August 2018; revised 2 July 2019; accepted 29 July 2019)

Abstract

Military applications of technologies for enhancing or producing vision play a key role in composing contemporary security, as such technologies are deployed to make security sense of everyday sociality, of battlefields, and of much in between these extremes. In this article, I set out to recompose militarised techno-vision through the public detritus left by its heterogenous development, use, and appropriation. I argue that as an heterogenous and amalgamated object, military techno-vision can be composed by speaking the stories of its leftovers, and that this composition is characterised by and in turn characterises a longstanding dilemma between fact and vision – between the ambiguity that is constitutive of the human practices of visual perception and image-making, and the desire for machines that can produce visual ‘actionable intelligence’ that can underpin security decisions. Discourses, practices, and regimes of visibility are deployed alongside technologies to occlude the ambiguity of technological vision and sustain the imaginary of technologically altered vision as neutral production of military or security facts.

Keywords: Visual Security; Critical Security Studies; Machine Vision; Combat Video; Art Photography; Posthuman Semiotics

Introduction: Composing military techno-vision

What would the technologies employed for military ‘enhanced’ vision tell us if we could assemble them and question them about the work they are being asked to do for us? This is the question that I have long been interested in; it dawned upon me when reading Vinciane Despret’s fabulous animal stories of science-making peacocks and semiotic tics.¹ Like these animals, visual technologies participate in making sense of the world, including for the purposes of war and security, and participate in making (lethal) decisions and classifications, but – again like the animals used in the scientific studies that populate Despret’s work – they are rarely regarded as other than instruments of domination and fields of competition, rarely allowed to participate in speaking about the domains of politics, science, and technology they are part of making. Here, I hope, they are.

To some theorists of media interested in military matters such as Kittler, Virilio, and their many followers, the relationship between conflict and visual media technologies is close to inevitable, expressed, and even appearing to be determined in *c*, the constant denominating the speed of light and the highest speed with which to make information travel. With this idea comes the idea that vision is not just another sense/interface – which could be replaced by smell or hearing in differently configured technologies, but that optical media present unique opportunities for speed, and it becomes unsurprising from this perspective that in the military domain, forms of technological vision and visual technology have long played crucial roles in the assemblages

¹V. Despret, *What Would Animals Say if We Asked the Right Questions?*, trans. B. Buchanan (Minneapolis: University of Minnesota Press, 2016).

that compose and manage global politics.² From less technocentric or – deterministic points of view, ethical desires to ground security decisions in facts or regimes of precision and traceability³ can lead towards similar appreciations for technological forms of vision that leverage the epistemic authority of images,⁴ yet in both the goal that technologies are employed to serve easily becomes privileged *vis-à-vis* the specific technologies serving them. By telling the stories of the detritus that is left over from the development and use of military techno-vision, I assert that, when told together, the stories of these leftovers can speak to us about the kind of desires, agents, and logics that populate the field, in a way similar to how Despret's fabulous animal stories speak of posthuman agencies and ethics,⁵ and thus participate in re-composing military techno-vision as an object.

I therefore ask in this article how we can make sense of the almost inaccessible sensemaking activities taking place in the field of military techno-vision, and propose that the answer can lie in speaking to the objects through which we encounter military techno-vision. One way of doing so, I argue, is to assemble military techno-vision as an amalgamated and composite object of concern in which the parts encountered – despite the heterogeneity of its constituent practices, technologies, and discourses – have some common attributes and are, as I will show, asked to perform common epistemic operations. The objects encountered also, I try to show, subtly resist or at least create some friction between themselves and the discourses of those deploying them. Visual technologies do not merely facilitate the goal of making 'near real-time'⁶ sense of sociality from a distance, but in doing so they also perform occasionally lethal categorisation, interpretation, and decision in relation to the world they interact with, which is both why such technologies are of concern to me here and why friction is important between technology and the discourse it is deployed with. I examine how the detritus left by these technologies speaks of such resistance or friction, enabling a questioning of the work we ask these technologies to do. In doing so, I am arguing that there is value in not attempting to gain privileged access to shielded technologies, but instead that we can gain unique insights by recomposing military techno-vision from the heterogeneous detritus its development use, and appropriation has left behind in the public space.

I start here by setting out why it is important to try to access deliberately opaque technologies in this way, thinking about how the performativity of different modes of inquiry may differ and by introducing the three stories that I tell. Thereafter, I briefly introduce how this method of composing an absent object relates to previous aesthetic inquiry in IR. In the following three sections, I then tell the stories of unruly evidence produced by gun camera sights, unanticipated semiotics revealed during machine vision development, and the ambiguous beauty produced by disused photographic film that ended up in the camera of an artist portraying eastern Congo, before concluding with thinking about how the stories told encounter similar themes revolving around the epistemic politics of leveraging visual technologies for purposes of security and warfare.

²J. Der Derian, *Virtuous War: Mapping the Military-Industrial-Media-Entertainment Network*, 2nd edn (London: Routledge, 2009); R. Bishop and J. Phillips, *Modernist Avant-Garde Aesthetics and Contemporary Military Technology: Technicities of Perception* (Edinburgh: Edinburgh University Press, 2010); R. Stahl, *Through the Crosshairs: War, Visual Culture, and the Weaponized Gaze* (New Brunswick, NJ: Rutgers University Press, 2018).

³As discussed in J. P. Burgess, *The Ethical Subject of Security: Geopolitical Reason and the Threat against Europe* (London: Routledge, 2011); M. Zehfuss, 'Targeting: Precision and the production of ethics', *European Journal of International Relations*, 17:3 (2011), pp. 543–66.

⁴For a discussion of visual epistemic authority in relation to security, see R. Saugmann Andersen, 'Epistemic authority, lies, and video: the constitution of knowledge and (in)security in the video/security nexus', *JOMEC Journal*, 4 (2013); R. Saugmann Andersen, 'Video, algorithms, and security: How digital video platforms produce post-sovereign security articulations', *Security Dialogue*, 48:4 (2017), pp. 354–72.

⁵Despret, *What Would Animals Say if We Asked the Right Questions?*

⁶The quotidian military concept of 'near real time', I think, expresses well the practical discourse that privileges optical media in contemporary military practice. Its centrality to practical military thinking was highlighted to me in a recent interview with Danish Air Force personnel. See K. L. Jacobsen and R. Saugmann, 'Optimizing coalition air warfare: the emergence and ethical dilemmas of red card holder teams', *Global Policy* (2019).

Detritus, black boxes, and non-mimetic composition

In contrast to the superhuman clarity allegedly provided by militarised techno-vision, the systems underpinning different practices of more-than-human vision are most often kept opaque and inaccessible for the publics who are asked to believe in but not allowed to peer through advanced scopes, or to scrutinise the lines of code through which visual technologies of surveillance and lethality enact, compose, and engage the world.

When seeking to make sense of, or with, technology, and likewise of the military use of technology, questions of secrecy, competence, and access are likely to appear as are issues of official and commercially motivated secrecy and problems caused by innate technical opacities. Military technologies are, however, far from the only area of social life in which such conditions are prevalent. Rather, as Frank Pasquale finds, such conditions are indicative of what he terms ‘the black box society’, an increasingly prevalent mode of governing that it is not driven by technological sophistication per se but in which considerations about the appropriate mix of secrecy and transparency are giving increasing priority to secrecy. Strategies combining ‘real secrecy’ – like locked doors and passwords – with ‘legal secrecy’ (obligations not to share knowledge), and ‘obfuscation’ about the workings of advanced technologies are adept at creating black boxes of decision-making closed off from citizen scrutiny,⁷ with security rationales providing what Pasquale terms a ‘formidable legal armamentarium’⁸ for these practices. While in institutional setups such as patent law, secrecy was balanced with transparency towards a trusted agency, today’s technologies are kept far more opaque. The ‘move from legitimisation-via-transparency to protection-via-secrecy was the soil of which the black box society sprang’.⁹

The unearthing of black-boxing as a performative governance logic rather than an unfortunate circumstance arising from complexity as well as military and commercial competition pressures is inspirational in that it invites both reconsidering the idea of a black box, its performativity, and the response to it. Black-boxing logics at work around military techno-vision create conditions in which the promotional narratives of actors engaged in the area encounter fertile ground for creating public perception of the area. Yet even if obfuscation and secrecy abound, both specific visual technologies and the military application of them do exist in relation to publics, and will rarely be completely shielded from these. Rather than seek privileged access and engage in the ‘implicit bargaining’ that come with such efforts,¹⁰ what I set out to do in this article is to engage in composing military techno-vision as an object of concern via the detritus left in the public space by different practices such as development, military use, and scrutiny or appropriation of objects that can be assembled as parts of a composite object of concern. Composing, in the sense described by Bruno Latour, is ‘not too far from “compromise” and “compromising”,¹¹ and indeed what I try to achieve in this composition is both to compromise by juxtaposing different heterogeneous leftovers without glossing over their heterogeneity and to be compromising by using the public detritus from a field of action to speak about the instabilities and fragility that is glossed over by black-boxing, secrecy, and obfuscation.

It is paramount here that this composition is not mimetic in the sense that I hold military techno-vision to be an object that exists, fully fledged and ready for critical interrogation in a reality external to this treatment – that is, as possessing the elements of anteriority, independence, and definiteness that John Law¹² treats as questionable parts of ‘out-thereeness’. Rather, military

⁷F. Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information* (Cambridge, MA: Harvard University Press, 2015), 6.

⁸*Ibid.*, p. 193.

⁹*Ibid.*

¹⁰T. A. Undheim, ‘Getting connected: How sociologists can access the high tech elite’, *The Qualitative Report*, 8:1 (2003), pp. 104–28 (p. 122); for the military field, see also E. Ben-Ari and Y. Levy, ‘Getting access to the field: Insider/outsider perspectives’, *Routledge Handbook of Research Methods in Military Studies* (Milton Park: Routledge, 2014), pp. 9–19.

¹¹B. Latour, ‘An attempt at a “compositionist manifesto”’, *New Literary History*, 41 (2010), pp. 471–90 (p. 474).

¹²John Law, *After Method: Mess in Social Science Research* (London: Routledge, 2004), p. 24.

techno-vision is ‘built from utterly heterogeneous parts that will never make a whole, but at best a fragile, revisable, and diverse composite’.¹³ It is my concern with engineering vision and applying the results of such engineering to security and conflict that drives me to assemble and label *military techno-vision* from the detritus left in the public sphere by such engineering and use. Like all composites it may crack or dissolve when exposed to heat or stress, or it may fail to amalgamate and remain separate heterogeneous parts.

Three fables about military techno-vision

Ideas and stories that may help to identify what military techno-vision might be can be sourced from the different encounters between visual technologies, the militarisation of these, and the public spheres they interact with. From the inception of a technology that usually takes place in an at least partly civilian engineering lab, through its classified use by the professionals of security, to its possible afterlife as a once-again civilian technology, black-boxing will never be complete and bits and pieces of detritus will be left in the public space. I choose here to pick up detritus that originate from the civilian development of technologies that are simultaneously undergoing militarisation, from the application of visual technologies in warfare, and from the civilian repurposing of once-militarised visual technologies. Inspired by Despret’s fables that permit ‘understanding how difficult it is to figure out what animals are up to’¹⁴ without reducing their diversity, I relate the stories enabled by these pieces of technology-related detritus, asking what it is these technologies are up to, and – adding considerations that relate to black-boxed technologies rather than to animals, and to the ever-pertinent question of how humans and technologies reconfigure the agency of each other¹⁵ – what they think of how we think about them and interact with them. In contrast to Despret’s diverse fables that do not aim to characterise animals but rather to argue in favour of not reducing them, I compose an object of concern out of the stories I tell, inevitably performing a somewhat reductionist move but doing so to be able to speak of how we interact with these different technologies in similar ways. I thus inquire into how human and technological agency is reconfigured around the object of military techno-vision, in technological practice as well as in the discourses surrounding technologies, thinking the composite of military techno-vision without glossing over or reducing the heterogeneity of the phenomena within it. With this I intend to enable a conversation about the operations performed with military techno-vision, and how black boxes, discourses, and technical practices relate to this.

The detritus I pick up for this effort is pieces of published or leaked footage, scientific controversies, and an artistic engagement with a declassified technology of militarised vision.

First, I tell of two relatively common access points through which we may be able to peer though military techno-vision. I consider the leaked battlefield video published by WikiLeaks with the title ‘collateral murder’, and relate it to a promotional video of precision strikes, published by the Israel Defense Forces, an actor that pioneered this now widespread practice.

Second, I pick up pieces from public engineering debates over a technology that is currently at the forefront of military efforts to leverage new visual technologies, machine vision technology that is able to autonomously perform object and pattern recognition tasks. Similar technologies are being put to use by the US Department of Defense to perform drone image analysis, and so inquiring into debates over testing image databases and the problems of reliability that emerge therein reveals much about both the actual workings of these complex systems and about the parts we play in their operation.

¹³Latour, ‘An attempt at a “compositionist manifesto”’, p. 474.

¹⁴B. Latour, ‘Foreword: the scientific fables of an empirical La Fontaine’, in Despret, *What Would Animals Say if We Asked the Right Questions?*, pp. vii–xv, vii.

¹⁵L. A. Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions*, 2nd edn (Cambridge: Cambridge University Press, 2007).

Third, I turn to the artist Richard Mosse's photographic project 'Come Out (1966)', in which he uses a disused military film, designed to reveal what is invisible to the unassisted eye, to portray the conflict in eastern Congo, a conflict that is for Western spectators particularly hard to immediately visually classify in terms of 'good' and 'bad' actors, highlighting the work done by such classification.

Military vision technologies and regimes of visibility

Seeing, even if grounded in the biological affordances of our species, is a thoroughly social process. It is full of expectations, desires, taboos, learned associations, and rules.¹⁶ The grids of intelligibility formed by these rules have been theorised under many different names that cover what I see to be roughly similar ideas. They are variously called 'scopic regimes',¹⁷ ways of seeing,¹⁸ aesthetic regimes,¹⁹ or regimes of visibility (and vision)²⁰ or, in an IR context, 'regimes of perceptibility',²¹ epistemic regimes,²² or scopic regimes.²³ Common among these ideas is that they describe how images enact divisions of the sensible,²⁴ forming regimes that tell us how to make sense of visual information and what to consider meaningful.

Pictures and the technologies and devices producing them are inevitably agents of such regimes, which are not only constituted around vision but around posthuman reconfigurations of biology, technology, and culture. Film and photography is deployed by militaries and security agencies to know society in superior ways, isolating and dissecting societal behavioural patterns and making visible neglected details, providing some form of advantage over an enemy or, increasingly, revealing what is held to be possible or precursory forms of enmity where none are obvious.

At the most basic level what any image does is highlight the transformation effected by the photographic image – the turning of a lived experience into an object²⁵ that can be looked at and stored but also circulated, mass reproduced, tampered with, or destroyed. This objectification is often held to have come into its own in Western visual culture through seventeenth-century Dutch painters such as Vermeer,²⁶ who drew on renaissance traditions of perspectivalism but exceeded and fractured these to render narratives – which were the focus of previous visual and often religious art – opaque, while highlighting objects and their realistic representation.²⁷ This Western pictorial tradition, coming from the oil painting as it was developed in Dutch art by building upon some of the inventions of the renaissance such as the Albertian and Brunelleschian forms of perspective, 'defines the real as that which you can put your hands on'.²⁸

Recognising that contemporary militarised seeing is dependent on rendering sociality recordable or storable in image objects that can be traced, mapped, stored, and analysed is thus not only about recognising a technological reality, it is also inescapably about recognising a cultural project. Michael Shapiro points to this, without problematising the object quality of images, when he

¹⁶M. Jay, 'Scopic regimes of modernity', in H. Foster (ed.), *Vision and Visuality* (Seattle: Bay Press, 1988).

¹⁷Jay, 'Scopic regimes of modernity', drawing on Christian Metz.

¹⁸J. Berger, *Ways of Seeing* (London: British Broadcasting Association and Penguin, 1972).

¹⁹J. Rancière, *The Politics of Aesthetics: The Distribution of the Sensible* (New York: Continuum, 2004), p. 8.

²⁰H. Foster, *Vision and Visuality* (Winnipeg: Bay Press, 1988).

²¹R. van Munster and C. Sylvest, 'Documenting international relations: Documentary film and the creative arrangement of perceptibility', *International Studies Perspectives*, 16:3 (2015), pp. 229–45.

²²R. S. Andersen, 'Remediating Security: A Semiotic Framework for Analyzing How Video Speaks Security', University of Copenhagen, Department of Political Science (2015), pp. 187ff.

²³K. Grayson and J. Mawdsley, 'Scopic regimes and the visual turn in International Relations: Seeing world politics through the drone', *European Journal of International Relations*, 25:3 (2018), available at: {doi: 10.1177/1354066118781955}.

²⁴Rancière, *The Politics of Aesthetics*.

²⁵cf. Berger, *Ways of Seeing*, p. 47.

²⁶Jay, 'Scopic regimes of modernity', p. 12.

²⁷Rancière, *The Politics of Aesthetics*, p. 24.

²⁸Berger, *Ways of Seeing*, p. 88.

notes that ‘the sense derived from seen objects is not merely a function of the degree of optical resolution; it derives from the projects and culturally induced expectations of the observer’.²⁹ This ‘sense derived’ refers precisely to the semiotic process of interpretation that images are subjected to and engaged in.

Yet images are not simply objects that store reality unaffected, until we the viewers bring our cultural expectations to them. Engaging technological forms of vision developed in the last decades of the twentieth century, and taking the engagement with technology as a departure for thinking about the agency of all forms of seeing, Donna Haraway concludes that ‘the “eyes” made available in modern technological sciences shatter any idea of passive vision; these prosthetic devices show us that all eyes, including our own organic ones, are active perceptual systems, building in translations and specific *ways of seeing*’.³⁰ These two engagements, Haraway’s and Shapiro’s, go in opposite directions in the process of seeing, yet with similar aims. While Haraway sets off on a technoscientific project pointing to the transformative agency of recording, Shapiro proceeds in a sociological direction and does the same with respect to the interpretative act of viewing. What opens then, for me, the most interesting perspective here is the fusion of these two directions not in people that see and interpret but in technological systems that not only record, store, and transmit images for people to interpret, but make image interpretation an automated or autonomous technological process too, fusing Shapiro’s ‘observer’, bearer of culturally induced expectations, with Haraway’s active perceptual prosthetic devices. Thinking with scopic regimes as part of composing military techno-vision thus emphasises the definition of the real enacted through the active translation of seeing into objects and through the grids of intelligibility that govern sensemaking.

Air war footage and the fogs of war

In contemporary warfare, video feeds often sustain Western military operations in the distant battlefields of the ‘everywhere’ war, contributing to keeping Western casualties low while projecting violence in a low average intensity across vast areas of the world. Video technologies permit the overlay of biological vision with heat-sensing, camouflage-detecting, night-vision, and other technologies of imagination that facilitate military advantage, and permit ‘near real time’ command and control over violence from afar. As most of these technologies automatically store the resulting images, they also create a visual archive of the violence they deploy. But despite countless hours of video of ‘close air’, ‘dynamic’ combat or pre-planned targeting, relatively little makes its way to the eyes of Western publics in the name of whose security the conflicts take place. Most videos seem to be amateur *milblog* compilations remediating dehumanised ‘target’ strikes, decontextualised to the point of being virtually meaningless.³¹ Some air war footage is published online through mainstream news sources, for example, the Associated Press’s ‘Raw Video’ series, giving an intelligible but promotional rendering of the battlefield operations they portray,³² as these videos are presumably published with official consent.

In this context, the US gun camera video leaked from Iraq by Chelsea Manning and published by WikiLeaks as ‘Collateral Murder’ functions as a meta-image,³³ as it becomes a rare piece of public detritus that stands in for a vast genre hidden from public scrutiny.

²⁹M. J. Shapiro, *Cinematic Geopolitics* (London: Routledge, 2009), p. 66.

³⁰Donna Haraway, *Simians, Cyborgs, and Women: The Reinvention of Nature* (London: Free Association Books, 1991), p. 190.

³¹Any online video search involving the keywords ‘strike’ and ‘insurgents’ will demonstrate this.

³²While this is the general trend, both declassified and leaked military video can be found online. See, for example, ‘rTrThanos’, *A-10 CAS* (2012) [unclassified]; and The Sun and ‘latestfacts’, *Friendly Fire Cockpit Video Iraq 2003. Matty Hull Killed* (2007) [leaked].

³³W. J. T. Mitchell, *Picture Theory: Essays on Verbal and Visual Representation* (Chicago: University of Chicago Press, 1994).

On 5 April 2010, the whistleblower organisation WikiLeaks released a video showing some forty minutes of classified gun camera footage leaked to it from the battlefield in Iraq, publishing it under the loaded title ‘Collateral Murder’. According to the organisation – and soon confirmed by US military sources – the video showed a 12 July 2007 series of air strikes taking place in Baghdad, in which two Reuters employees and ‘over a dozen’³⁴ unidentified Iraqis were killed. Alongside the forty-minute ‘raw’ version (containing the original decrypted video with only subtitles added) an edited 18-minute version with subtitles, explanatory text superimposed onto the images and a 2.25-minute introduction to the events depicted was released. All releases were made on a tailor-made website, *collateralmurder.org*. The website featured both versions of the video, embedded from YouTube; captioned stills; a written summary overview of the video as well as the also leaked US Rules of Engagement from around 2007; biographical material and family testimonies relating to the two Reuters employees who were reportedly victims of the shooting seen in the video, and a lot of other background material.

These (image)texts and contexts were immediately seized on by news media, political professionals, commentators, and bloggers and remediated in various formats, from television broadcasts to tweets. By the time of writing the two versions of the video have been watched more than nineteen million times online, with the short, edited, and commented version accounting for the vast majority of those views, more than sixteen million, and the longer accounting for 2.8 million views.³⁵

The video appears as a piece of scrap evidence, apparently recorded by the gun-mounted camera of the helicopter gunship involved in the incident, and gives access to a hitherto obscure reality – claiming that *this is how it was*³⁶ in the obscured battlefield of Baghdad. Crucially, it also claims an aesthetic immediacy,³⁷ that *this is how it was seen* – seen by the US military apparatus involved. This indexical-iconic reading of the visual signifiers as causally connected to and resembling the signified³⁸ is reinforced by military technoscientific discourses – on surgical strikes, revolutions in military affairs and the like – that claim (surgical) accuracy and superhuman (enhanced) vision. WikiLeaks’s release aimed to turn the results of this discursive work³⁹ (performed to legitimise militarised technology) from being an instrument of military propaganda towards being a vehicle of critique, yet as I will show in the following section, the video did not succeed in questioning militarised vision because, rather than being allowed to speak, it was deflected by US authorities and weakened by the actions taken by WikiLeaks itself to promote and direct our viewing of the video.

Occluding evidence: the performance of a semiotic fog of war by US authorities

As the video spread, US defence authorities were called upon to explain how such a seemingly incriminating video could exist, documenting a well-known encounter that had been the object of public controversy (as two Reuters staff were killed) and of which no documentation was said to be available following Reuters’ earlier inquiries. The US authorities countered the video with two interrelated arguments, one about the partiality of the view offered and another deploying the metaphor of ‘the fog of war’ to argue that seeing the evidence contained in the video did not

³⁴This figure is from the presentation at: {*collateralmurder.com*} accessed 14 September 2012.

³⁵YouTube viewcounts as of 3 August 2018: Long version 2,844,019; short version 16,307,482. Of course there are also innumerable clips, identical versions posted at different channels, shorter versions, commented versions, and analyses, etc. These are not included in this count, which is thus very conservative in terms of assessing the online spectatorship of the leaked video. The television show *The Young Turks*’ analysis, for example, has been viewed more than 1.3 million times.

³⁶R. Barthes, ‘Rhetoric of the image’, *Image Music Text* (London: Fontana Press, 1977), pp. 32–52 (p. 44).

³⁷J. Bolter and R. A. Grusin, *Remediation: Understanding New Media* (Cambridge, MA: MIT Press, 2000), p. 21.

³⁸C. S. Peirce, *Peirce on Signs: Writings on Semiotic* (Chapel Hill, NC: University of North Carolina Press, 1991).

³⁹T. Gillespie, ‘The relevance of algorithms’, in T. Gillespie, P. J. Boczkowski, and K. A. Foot (eds), *Media Technologies: Essays on Communication, Materiality, and Society* (Cambridge, MA: The MIT Press, 2014), pp. 167–93 (p. 188).

equate seeing what happened in the lethal encounter. The ‘*lack of context*’ argument holds that when watching the forty minutes of videotape containing the view from one gun camera as well as the crew radio, one should not confuse vision and knowledge: “You are looking at the war through a soda straw and you have no context or perspective”⁴⁰ this argument holds. ‘One cannot judge it from the video alone’, asserted an analyst,⁴¹ echoing the comments by the US Secretary of Defense that ‘there’s no before and there’s no after ... These people can put out anything they want.’⁴² This argument attacks the epistemic strength of the leaked video as a source of knowledge, challenging its position as a ‘witness’ able to give evidence. Second, the ‘fog of war’ metaphor extends the lack of vision argument from the video to the general situation, with the US Secretary of Defense asserting that ‘you talked about the fog of war. These people were operating in split-second situations.’⁴³ The narrative locates responsibility for any mistakes with the pressure of combat and laments that the reality of war is difficult, dismissing the technological devices providing enhanced vision. This renders the video hardly intelligible, but rather than attacking its epistemic authority as a documentation of war, it seems to extend this attack to the devices deployed in war, rendering the visual technologies of lethality unfortunately inadequate for penetrating the mythical fog of war. Choosing not to take responsibility for the careless deployment of lethal techno-vision at or beyond the very limits of its intelligibility – where all that can be seen and all that matters is bodies that, simply by being and living in the midst of what the US considered a battlefield, are rendered expendable and ungrievable – Gates deploys the fog of war narrative to imply that US forces do not see what they shoot very well, and uses it as a defence against accusations of carelessness. ‘It’s painful to see, especially when you learn after the fact what was going on’,⁴⁴ implying that the technological vision deployed is not adequate for knowing what is going on. Two observations are interesting. First, the pilots indicated in sworn statements taken by the army investigation that ‘[v]isibility was good. Once we started the engagements of course, it kicked up dust so you know, it’s harder to see’,⁴⁵ so the deployment of the fog of war seems not come from the situation. Second, no individual was faulted in any of the official investigations.⁴⁶ Yet in accepting ‘Collateral Murder’ as real, acknowledging its aesthetic immediacy, but reconfiguring the technology involved as unable to provide reliable knowledge, Gates enacts through the kind of realism described by Rancière as stemming from modern literature’s ‘fragmented or proximate mode of focalization, which imposes raw presence to the detriment of the rational sequences of the story’.⁴⁷ Here, the limits of hypermediated⁴⁸ technologically enhanced visibility are magnified and enacted as a condition (the fog of war) rather than a choice based on policy (a practical weighing of how well soldiers need to see what they target according to rules of engagement; official responses to previous possibly unlawful killings; etc.). Gates thus sustains that the

⁴⁰US Defense Secretary Gates, cited in S. Allan and K. Andén-Papadopoulos, “‘Come on, let us shoot!’: WikiLeaks and the cultures of militarization”, *Topia: Canadian Journal of Cultural Studies*, 23–24 (2010), pp. 244–253 (p. 250).

⁴¹From a Danish defence think tank in Danish daily *Politiken*: ‘Krigsvideo fører til krav om undersøgelse’ (2010). For an almost identical statement from a US Heritage Foundation analyst, see ‘Anklager: USA nedskød civile i Irak’, *Politiken* (2010).

⁴²As cited in Reuters World Service, ‘Gates Assails Internet Group Over Attack Video’ (13 April 2010), available at: {<https://www.reuters.com/article/us-iraq-usa-journalists/gates-assails-internet-group-over-attack-video-idUSTR63C53M20100413>}.

⁴³Ibid.

⁴⁴US DoD, ‘News Transcript ABC’s “This Week” Interview with Secretary Gates and Secretary Clinton Presenters: Secretary of Defense Robert M. Gates and Secretary of State Hillary Clinton April 11, 2010’ (2010).

⁴⁵15-6 Investigation Officer, ‘Sworn Statements for 1st Air Cavalry Brigade 15-6 Investigation 1 ACB (Reuters) _12_JUL_07’ (2007), p. 13.

⁴⁶M. A. Hasian, ‘Watching the domestication of the Wikileaks helicopter controversy’, *Communication Quarterly*, 60:2 (2012), pp. 190–209 (p. 198).

⁴⁷Rancière, *The Politics of Aesthetics*, p. 24.

⁴⁸Bolter and Grusin, *Remediation*.

ground truth is both unknowable from the decontextualised video evidence and that in context the video technologies employed were not able to penetrate the fog of war.

This chain of arguments can be understood as playing with two aesthetic concepts of scholarship on mediation. First, Gates mobilises hypermediacy – the ‘windowed’ aesthetic of media-within-media⁴⁹ – to make the *lack of context* argument that renders ‘Collateral Murder’ unimportant and unreliable. Gates’s acknowledgement of the immediacy of the evidence is supplemented by his denunciation of it and his rendering of our seeing as hypermediated – as a leisurely form of spectatorship that is different from the immediacy of the battlefield. Hasian finds this idea twisted to one of professionalised vision in the ‘many milblog commentaries that underscored the point that only those who had guarded democratic nations and survived the crucible of war had the ability or right to authoritatively comment on the “context” for interpreting the Apache video feeds’.⁵⁰ This defence enacts a militarised critique of civilian spectatorship as unreliable qua hypermediated and civilian, the hypermediation serving to deprive the very images used by soldiers in targeting of their power to know the battlefield.

Along similar lines, Mette Mortensen⁵¹ takes the instability of meaning derived from the video (the mistaking of a camera for an RPG, for example) to question the very notion of ‘visual evidence’. She contends that ‘the reliability and readability of the gun camera tape are questioned’ by the mistakes made in the video. Yet in their attempts to weaken the epistemic status of the video, US military authorities attack the lack of context and the ‘fog’ occluded by this soda straw view, rather than the footage itself. The video is taken to mean different things by different actors and in different remediations, but its lack of reliability as witness is not due to its quality, but rather due to lack of context or military training, suggesting that military techno-vision is a particular human-machine configuration that cannot be replicated in public. While acknowledging that the video gave the crew the intelligence it needed for acting, the US military establishment seeks to herd the image back into the fold of impermeable ambiguity and contain the epistemic authority to produce security knowledge from the video sequences within the spheres of the ‘professionals of security and politics’.⁵² But the authorities responsible for the violence shown in the video are not the only ones who seek to influence the video and direct who can interpret it and how. This ambition is shared by the publisher of the leaked video, WikiLeaks.

In its seeking to control the interpretation and spread of the video, WikiLeaks deploys different semiotic techniques, namely anchoring and editing.⁵³ Drawing on dominant representational codes of mediating remote warfare, it overlays the video with explanatory textual anchors and uses editing to create a spectacular short version of the video in which a clear narrative – of careless US military staff – stands out. In doing so, WikiLeaks manipulates the video to fit what Lilie Chouliaraki terms “ecstatic news” – news so extraordinary that it warrants live coverage beyond the normal news bulletin, bringing global audiences together around a 24/7 mode of reporting⁵⁴ – furthering the remediation of the video. But this spectacular editing also keeps the video firmly *within* a traditional set of representational codes, seeking to utilise these for partly differing purposes: while one common purpose is to control interpretation and maximise spread of a video, Wikileaks seeks to criticise rather than celebrate or motivate the depicted acts of war. In its desire to produce effortless ecstatic news – a ‘corroborative’ spectacle rather than ‘pensive’ video that would cause the spectator to

⁴⁹Ibid., p.12.

⁵⁰Hasian, ‘Watching the domestication of the Wikileaks helicopter controversy’, p. 191.

⁵¹Mette, Mortensen, ‘Metacoverage taking the place of coverage: WikiLeaks as a source for production of news in the digital age’, *Northern Lights*, 10 (2012), pp. 91–106 (pp. 98–9).

⁵²D. Bigo, ‘Security, exception, ban and surveillance’, *Theorizing Surveillance* (Devon, UK: Willian Publishing, 2006).

⁵³For more on these techniques, see Andersen, ‘Remediating Security’, p. 185 (anchoring), p. 224 (editing).

⁵⁴L. Chouliaraki, ‘Re-mediation, inter-mediation, trans-mediation: the cosmopolitan trajectories of convergent journalism’, *Journalism Studies*, 14:2 (2013), pp. 267–83, 6.



Figure 1. Screenshots from Israel Defense Forces video⁵⁷ (left) and from the WikiLeaks-edited version of ‘Collateral Murder’⁵⁸ (right).

‘see’ and think, but also provide room for interpretation⁵⁵ – WikiLeaks is compromising its critical ability, denying the video the room to speak for itself. This entails a denial of visual ambiguity in favour of an easily recognisable visual narrative with good and bad actors.⁵⁶

Figure 1 illustrates how the anchoring techniques employed by Wikileaks are shared with the promotional videos published by the Israel Defense Forces (IDF), an actor at the forefront of using video in propaganda efforts. These come in many formats, but a particular type or genre is videos that ‘reveal’ the presence of Hamas’s deceptive military installations in civilian areas and show the destruction of these installations by precision strikes. The typical example has footage from a drone or other air surveillance, overlaid with text anchors exactly like the ones used by Wikileaks but labelling this or that building as a ‘Hamas rocket factory’ or ‘launch site’, seconds before the facility is destroyed by a missile.⁵⁹

Using text anchors imposes sense onto the images and forces distinctions onto the otherwise indistinguishable persons and buildings seen, in the case of IDF rendering them less-than-civilian and thus ungrievable.⁶⁰ To Neve Gordon and Nicola Perugini, such terms applied to persons in a battlespace ‘actively participated in structuring acts of war’,⁶¹ functioning as speech acts that have devastating legal and practical consequences. In their analysis of the deployment of the term human shield they note how, ‘[a]s a perlocutionary speech act, the term “human shields” bestowed a legal definition upon thousands of Iraqi civilians – before the assault on Mosul even began – that preemptively relaxed the conditions under which Iraqi forces and their allies could deploy violence.’⁶² The efficiency of disciplining images via overlaid ‘explanatory’ text here lies in its giving purpose to what would otherwise appear as the unmotivated bombing of

⁵⁵M. J. Shapiro, *Studies in Trans-Disciplinary Method: After the Aesthetic Turn* (London: Routledge, 2013), p. 24.

⁵⁶For a discussion of this issue, see R. S. Andersen and F. Möller, ‘Engaging the limits of visibility’, *Security Dialogue*, 44:3 (2013), pp. 203–21.

⁵⁷Israel Defense Forces, ‘Pinpoint Targeting of a Hamas Weapon Manufacturing Facility’ (2012).

⁵⁸WikiLeaks, ‘Collateral Murder’ (short version) (2010).

⁵⁹The Israel Defense Forces have long been utilising this technique, and a great number of examples can be found on their YouTube channel, under the handle ‘idfnadesk’.

⁶⁰J. Butler, ‘Judith Butler – Frames of War: The Politics of Ungrievable Life’, public lecture presented by the Humanities and Arts Research Centre of Royal Holloway, the School of Psychosocial Studies (Birkbeck) and the Birkbeck Institute for Social Research (2009).

⁶¹Neve Gordon and Nicola Perugini, ‘Human shields, sovereign power, and the evisceration of the civilian’, *AJIL Unbound*, 110 (2016), pp. 329–34 (p. 330).

⁶²*Ibid.*, p. 300.

everyday infrastructure, such as houses and cars, by rendering these as targets, via claims that they are, for example, ‘bomb-storage sites disguised as greenhouses’.⁶³

This section has used air war video detritus – either leaked or published as part of public diplomacy efforts – to present my first fable of military techno-vision. The stories tell of how regimes of visibility are not only the result of secrecy around military techno-vision but also conditioned by complex human-technology reconfigurations such as those enacted in news media remediation criteria favouring visual representations that are both ecstatic and unambiguous.

The remediation regime described can be thought of a semiotic fog that abounds in the politico-epistemic landscape of video-mediated security and influences how military techno-vision is perceptible to the public. When you shed light on fog, you do not necessarily see the landscape any clearer. Rather, the light is likely to alert you to how you are immersed in fog and cannot see very far; and my aim here is indeed not to dissipate this fog but to tell of how it is composed and how it is rendered penetrable only by the professional vision of operators of technologically mediated modes of seeing that involve laser-guided missiles, heat-sensing and night vision scopes, and which obscure to the public the human vision of atrocities performed in plain view.

Through the practices of promotional publication of footage and news media remediation, we become accustomed to ‘our’ warring being visualised as precision bombing footage with anchoring labels – or, alternatively, rendered a distant spectacle of shock and awe. This makes it possible to assert that the view offered in leaked video is occluded by fog or by lack of information or professional viewing practices, rather than a *modus operandi* of air war where technologically enhanced vision is not able to do away with ambiguity. The normal ways of visualising war – where ambiguity is done away with through textual anchors and editing, and which WikiLeaks seeks to conform to in order to maximise remediation – thus operate as a semiotic fog, preventing us from making critical sense of what we see. Thus, ‘[w]hen we are able to note a marked failure of perception – cases of misrecognition of persons or objects – one can attribute it to the fogging effects of a productive consciousness rather than to a haze that exists in the world, intervening between understanding or vision and object.’⁶⁴ Rather than simply a haze, this productive regime of visibility is directional, steering interpretation of military footage towards reducing ambiguity and presenting techno-vision as full of actionable intelligence, and towards ecstatic representation that can sustain large media audiences. Thus, the *problématique* of military techno-vision and the public cannot be overcome by simply making images public. While the black-boxing of military techno-vision may in the first place enable discourses of surgical precision to coexist with large civilian losses, changing the regime of visibility is also a matter of changing the public structuration of vision.

Machines, vision, and security

After decades of development of ever-more powerful surveillance technology, the problem of visuality for security services is at the moment not sensing or optics, it is vision (or understanding). Terrorist incidents, such as the 2011 far-right attacks in Norway, leave authorities with tens of thousands of hours of relevant surveillance footage to review.⁶⁵ Already in 2011, ‘the US Air Force [had] amassed over 325,000 hours of drone video – that’s about 37 years of video’ and today a ‘single drone with these sensors produces many terabytes of data every day. Before AI [artificial intelligence] was incorporated into analysis of this data, it took a team of analysts working 24 hours a day to exploit only a fraction of one drone’s sensor data’.⁶⁶ These are not only data

⁶³From a prominent video on the IDF YouTube channel.

⁶⁴Shapiro, *Cinematic Geopolitics*, p. 64.

⁶⁵Norwegian Police Security Service (PST), ‘PST and Societal Security’ (2014).

⁶⁶G. C. Allen, ‘Bringing AI to the Fight against ISIS’ (21 December 2017), available at: <https://thebulletin.org/2017/12/project-maven-brings-ai-to-the-fight-against-isis/>.

problems, as already in 2010 the fog of war – the metaphor that is used to occlude erroneous lethal decisions – had become digital in the sense that the US had taken to blame drone strikes targeting civilians on digital ‘information overload’.⁶⁷ The contemporary problem is thus understanding, not information: ‘the defense intelligence community is currently drowning in data. Every day, US spy planes and satellites collect more raw data than the Defense Department could analyse even if its whole workforce spent their entire lives on it.’⁶⁸

Machine vision, the development of technology – often described as artificial intelligence and currently based on neural network computing architectures – to perform tasks like categorising content in images (for example, colours, shapes, or metadata) and searching for patterns in or between multiple images or matching image contents with an existing database, is the current proposition for turning these massive amounts of visual observation into the militarised version of understanding, actionable intelligence. Like the pieces of detritus from air warfare, there are also traces of machine vision in the public space, albeit not of exactly the same kind used in direct military applications. The traces of machine vision relate to software development controversies, from debates over the training and testing of software in image databases to debates over reliability and misrecognition. From these starting points, I work my way towards machine vision systems that are trained through them and the visual culture developed from databases and the algorithms working with them.

Training databases, ambiguity, and ground truth

A surprisingly small number of large image databases have been used for training and testing most of the leading machine vision systems built so far. This is due both to the cost of creating databases, the advantages of reusing that labour, as well as of the possibility for measuring different systems’ performance against the same databases. The most widely used training image database, ImageNet, is built by having search engines look for images following simple descriptors in major search engines, that is, it consists of images harvested online following, to some degree, the labels attached to them in search engines.⁶⁹ Search engine images reflect the cultures supplying these images as well as the terms used to find them, making them culturally specific. The constructors of ImageNet are aware of this in remarking that ‘[t]o further enlarge and diversify the candidate pool, we translate the queries into other languages, including Chinese, Spanish, Dutch and Italian’,⁷⁰ yet such enlargement also highlights the narrow language selection parameters being employed.

Images are described and categorised for ImageNet using the online labour-sourcing portal Mechanical Turk, where workers get paid in fractions of cents for each very small job they complete. After unknown workers produce and cross-verify each other’s descriptions, an image is entered into the database, which in 2014 ‘contain[ed] 14,197,122 annotated images organized by the semantic hierarchy of WordNet’.⁷¹ The database contains more than twenty thousand categories of objects, meaning that this is the number of object categories that machine vision systems can then be trained to recognise. The ImageNet database hosts an annual machine vision competition in which systems compete in *image classification* – which ‘tests the ability of an algorithm to name the objects present in the image, without necessarily localizing them’; *object detection and localisation* – which ‘evaluates the ability of an algorithm to localize one instance of an

⁶⁷T. Johnson and C. F. Wald, ‘To Handle Its Influx of Drone Footage, Military Should Teach AI to Watch TV’ (26 November 2017), available at: {<https://www.wired.com/story/the-military-should-teach-ai-to-watch-drone-footage/>}.

⁶⁸Allen, ‘Bringing AI to the Fight against ISIS’.

⁶⁹J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, ‘ImageNet: A Large-Scale Hierarchical Image Database’ (Miami, FL: IEEE, 2009), pp. 248–55 (p. 251).

⁷⁰Ibid., p. 251.

⁷¹O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, ‘ImageNet Large Scale Visual Recognition Challenge’, *arXiv:1409.0575 [cs]* (2014), pp. 7, 3.

object category'; and *object detection* – which evaluates the ability of an algorithm to name and localise all instances of all target objects present in an image.⁷² The competition is based on images harvested from Flickr⁷³ and has been credited with contributing to the recent acceleration in advances in deep learning following a breakthrough made in the 2012 contest.⁷⁴

The limitations of ImageNet are not only with regards to the finite number of categories and sources of image data. As the creators of a competing database built later, Microsoft Common Objects in Context (MS COCO), assert, 'current recognition systems perform fairly well on iconic views, but *struggle to recognize objects otherwise* – in the background, partially occluded, amid clutter – reflecting the composition of actual everyday scenes'.⁷⁵ This is the result of ImageNet (and other datasets') images being sourced primarily through search engine querying, producing 'iconic' images in which the sought-after object or attribute is paramount, leading to images that according to the MS COCO developers 'lack important contextual information and non-canonical viewpoints'.⁷⁶ In such settings, 'Object category presence is often ambiguous. Indeed ... even dedicated experts often disagree on object presence, e.g. due to inherent ambiguity in the image or disagreement about category definitions'.⁷⁷ Yet, as machine teaching goes, the dataset still works by assigning a 'ground truth', 'computed using majority vote of the experts'⁷⁸ to denote what the image 'truly' contains. Ambiguity and polysemy, which have been at the heart of thinking about seeing since at least Wittgenstein's duck-rabbit,⁷⁹ are thus eradicated in favour of a unitary ground truth in the database construction process and thus carried over into the interpretative work done by machine vision software.

Several other features of these datasets are interesting. Take, for example, an illustrative 'unsupervised' discovery, in which Google's machine vision system discovered a class of objects, and which serves to highlight the political economy of objects. The discovery in question involved 'a giant unsupervised [computer] learning system [that] was asked to look for common patterns in thousands of unlabeled YouTube videos'. After leaving the system to work, one of the researchers involved ushered over a colleague and said, "look at this". 'On the screen was a furry face, a pattern distilled from thousands of examples. The system had discovered cats.'⁸⁰ Anybody familiar with YouTube and the output of its recommendation algorithms⁸¹ will know that cats are at the apex of a YouTube's video-industrial complex, meaning that the discovery of cats also involved a feedback loop in which the computer system revealed the algorithmic governance of the visual public, rather than merely discovering furry animals.

Given the importance of and difficulty with datasets for developing machine vision systems, training data will likely also be an issue for surveillance and military applications of machine vision. Training datasets are thus, unsurprisingly, an issue in publicly known military applications of machine vision. The most well-known, Project Maven, a US Department of Defense project to establish an 'Algorithmic Warfare Cross-Function Team', was widely reported on in 2018 due to the revelation that Google was involved despite promises not to weaponise artificial intelligence, and the subsequent protests by Google employees that led to the company not renewing its contract and so terminating its involvement with the DoD.

⁷²Ibid., pp. 5, 8, 10.

⁷³Ibid., p. 4.

⁷⁴From not working to neural networking', *The Economist* (2016).

⁷⁵T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, 'Microsoft COCO: Common Objects in Context' (2015), p. 1, emphasis added.

⁷⁶Ibid., p. 4.

⁷⁷Ibid., p. 6.

⁷⁸Ibid., p. 6.

⁷⁹Mitchell, *Picture Theory*, p. 50.

⁸⁰From not working to neural networking', *The Economist*.

⁸¹On the term recommendation algorithm, see Saugmann Andersen, 'Video, algorithms, and security.'

Project Maven's objective is 'to turn the enormous volume of data available to the DoD into actionable intelligence and insights at speed', with the project's first tasks already familiar from above, namely to 'organize a data-labeling effort, and develop, acquire, and/or modify algorithms to accomplish key tasks'.⁸² Activists have warned that this effort is an important one on the road to autonomous weapons because it seeks to 'reduce the human factors burden [that is, labour needs] of FMV [Full-Motion Video] analysis, increase actionable intelligence, and enhance military decision-making'.⁸³ Project Maven thus 'seeks to automate basic labeling and analysis associated with full-motion video surveillance',⁸⁴ something that we shall see is far from unproblematic. It 'focuses on analysis of full-motion video data from tactical aerial drone platforms such as the ScanEagle and medium-altitude platforms such as the MQ-1C Gray Eagle and the MQ-9 Reaper',⁸⁵ some of the drones responsible for the current visual data deluge.

Even if Maven focuses on one class of images only, the differences in the size of image training data are striking: 'In Maven's case, humans had to individually label more than 150,000 images in order to establish the first training data sets; the group hopes to have 1 million images in the training data set by the end of January [2018]'.⁸⁶ Although these numbers seem impressive, we should recall that even ImageNet's fourteen million training images were described earlier as still only providing 'iconic' views of the objects identified in them. In comparative terms, even Project Maven's desired dataset is several orders of magnitude smaller than that of ImageNet, a somewhat alarming fact given its proposed lethal applications.

The current approach to dealing with this paucity of labelled data seems to be one of limiting the capabilities of the system – or, if we adopt the anthropomorphic lingo of artificial intelligence and machine vision, one of visual stupidity. Thus, Maven's 'immediate focus is 38 classes of objects that represent the kinds of things the department needs to detect, especially in the fight against the Islamic State of Iraq and Syria'.⁸⁷ While it may be desirable to start out with a narrow focus, such an approach obviously does not capture the ambiguity and multiplicity that is the 'ground truth' of those bearing the hardships of, for example, Daesh/IS and the fight against them, instead only sharpening the reductionism of the alternative machinic 'ground truth'.

Militarised machine vision will need training datasets that not only permit the systems learning from them to segment objects and distinguish between them but also be highly sensitive to context (think of a fighter with a gun versus a farmer with a gun), visual culture (for example, similar gestures with different cultural meanings), as well as enemies trying not to be classified as such.

Machine misrecognition

Seeking to get an idea of how machine vision works, and which controversies are debated within the community developing it, we eventually arrive at some of the most scrutinised systems, Google's publicly available image description and object identification systems. These, I thought, would have to stand in for the secret but probably similar systems that I suspected would be at work in military image databases, and this section was written on that background. Yet in the first months of 2018, it was revealed that Google was indeed providing machine vision services to the US Department of Defense, through Project Maven. The descriptions of this project as an effort that 'will 'provide computer vision algorithms for object detection, classification, and alerts for

⁸²Deputy Secretary of Defense, 'Establishment of an Algorithmic Warfare Cross-Functional Team (Project Maven)' (2017).

⁸³Ibid.

⁸⁴Johnson and Wald, 'To Handle Its Influx of Drone Footage, Military Should Teach AI to Watch TV'.

⁸⁵Allen, 'Bringing AI to the Fight against ISIS'.

⁸⁶Ibid.

⁸⁷C. Pellerin, 'Project Maven to Deploy Computer Algorithms to War Zone by Year's End' (21 July 2017), accessed at: {<http://www.defense.gov/Newsroom/News/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/>} (no longer online).

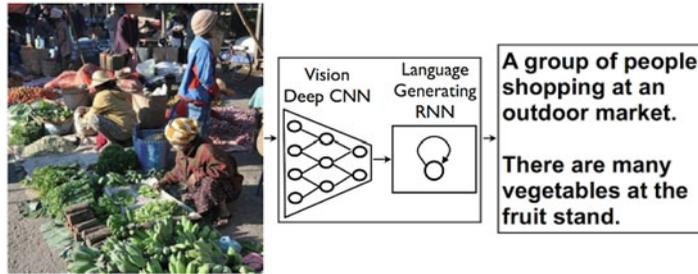


Figure 2. Representation of Google's image description and object identification system.⁸⁹

FMV [full-motion video] PED [processing, exploitation, and dissemination]’⁸⁸ brought uncanny memories of the features I had laughed at in descriptions of the image description and object labelling system, which are described below.

Google's image description and object identification system works through combining different neural networks, as shown in Figure 2, combining machine vision with natural language processing.

The use value of this combination – identification and description – lies partly in its making images compatible with other database objects, and easily searchable through the keywords and object descriptors generated. This epistemic rendering of the image rests, firstly, on an understanding of images as transparent or non-distorting – what can be termed an aesthetic of immediacy⁹⁰ in which iconic images, as described above, maintain an indexical connection to the world they depict. Secondly, it rests on an understanding of reality as limited and unambiguous. These properties are striking if we turn not to the research paper outlining the system but rather to the description of it on the Google AI Blog:

‘Two pizzas sitting on top of a stove top oven’, ‘A group of people shopping at an outdoor market’, ‘Best seats in the house’. People can summarize a complex scene in a few words without thinking twice. It's much more difficult for computers. But we've just gotten a bit closer – we've developed a machine-learning system that can automatically produce captions (like the three above) to accurately describe images the first time it sees them.⁹¹

The claims above are strikingly at odds with the understanding of images that one would get from the disciplines engaging images in the social sciences – visual studies, visual culture studies, visual semiotics, or the like. The sentences presented as analogous to how humans effortlessly describe an image do not sound like how most people would describe any scene or image, yet the creators assure that the system meets state-of-the-art thresholds and comes close to ‘human performance’ even with the added difficulty of formulating image captions in sentences.⁹² The idea that humans effortlessly describe an image in identical or like terms can be seen as the discursive upkeep that is always part of algorithmic governance,⁹³ in this case discursive work aiming at disciplining the images to fit with one of the fundamental ideas in machine

⁸⁸Deputy Secretary of Defense, ‘Establishment of an Algorithmic Warfare Cross-Functional Team (Project Maven)’.

⁸⁹O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, ‘Show and Tell: A Neural Image Caption Generator’, *arXiv:1411.4555v2* (2015), p. 1.

⁹⁰cf. Bolter and Grusin, *Remediation*.

⁹¹O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, ‘A Picture is Worth a Thousand (Coherent) Words: Building a Natural Description of Images’ (17 November 2014), available at: {<https://ai.googleblog.com/2014/11/a-picture-is-worth-thousand-coherent.html>}.

⁹²Vinyals et al., ‘Show and Tell’.

⁹³Gillespie, ‘The relevance of algorithms’, p. 188.

vision learning, the previously mentioned idea of a ‘ground truth’ that can be assigned to an image so that the machine vision system can learn to match this correctly with different images.

While the problematic idea of ground truth can also be interrogated through studying training database design, as above, what the debate about machine vision systems adds is not only whether or not images can be reduced to ground truths, but rather what is the agency of the translation of images that occurs when they become objects of machine vision, and the technological fragility in this transformation. In an intriguing article titled ‘Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images’,⁹⁴ the authors show that ‘it is easy to produce images that are completely unrecognizable to humans, but that state-of-the-art DNNs believe to be recognizable objects with 99.99% confidence’, some of which are presented in Figure 3. Using machine vision systems trained on ImageNet or similar large databases, and pairing these with evolutionary algorithms ‘that optimize images to generate high-confidence DNN predictions for each class in the dataset the DNN is trained on’,⁹⁵ the authors are able to supply machine

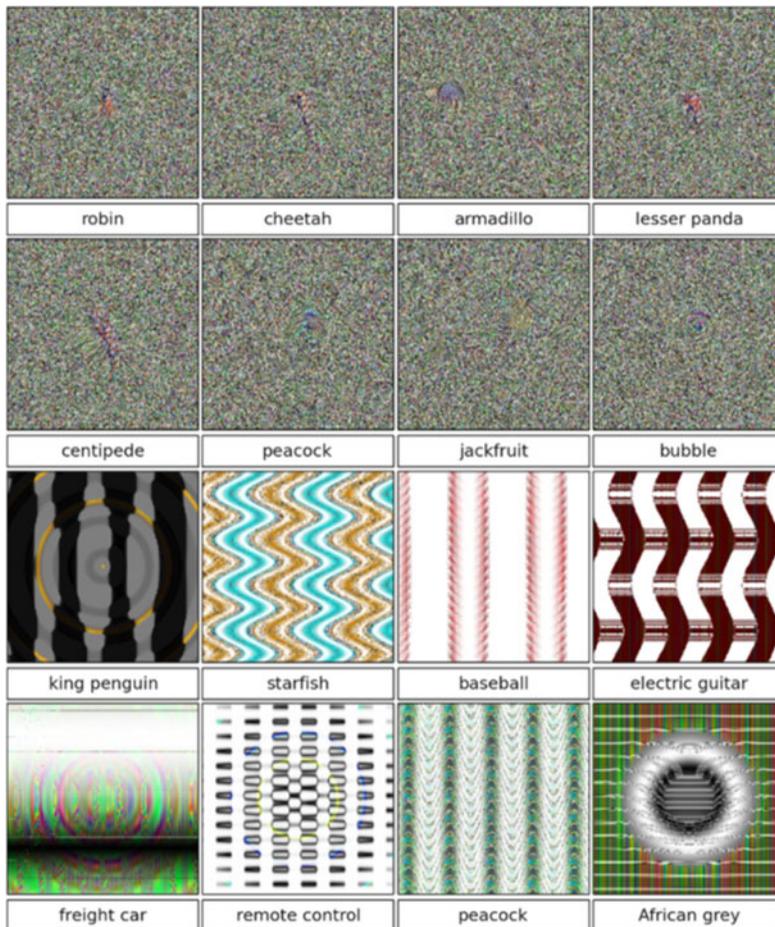


Figure 3. Fooling images and the labels they trigger. Illustration from Nguyen et al.⁹⁶

⁹⁴A. Nguyen, J. Yosinski, and J. Clune, ‘Deep neural networks are easily fooled: High confidence predictions for unrecognizable images’, *Computer Vision and Pattern Recognition* (2015).

⁹⁵Ibid.

⁹⁶Ibid.

vision algorithms with abstract or non-representational images and find images amongst these that are labelled by machine vision algorithms with very high levels of confidence. The technical process for doing so is not my concern here, rather I am interested in the character of the images, and the relations between images, discourses of machine vision, and the type of understanding sought or gained through machine vision.

The strangeness of these images that seem familiar to machine vision systems is intriguing, funny – and deeply uncanny given that Project Maven and probably other similar classified military projects are *already* employing machine vision to classify and follow objects of concern in the battlefield.⁹⁷ While the images in the lower panes betray some abstract symmetry and perhaps some flicker of likeness to that which they are mistaken as, preserving just a hint of the iconicity that pervades human visual meaning making, the images above show nothing that human understanding can relate to the categories they trigger in machine vision systems. Surprisingly, perhaps, the errors uncovered here seem to be rather robust, as ‘it is not easy to prevent the DNNs from being fooled by retraining them with fooling images labelled as such. While retrained DNNs learn to classify the negative examples as fooling images, a new batch of fooling images can be produced that fool these new networks, even after many retraining iterations.’⁹⁸

The detritus stemming from engineering controversies reveal the utopianism underpinning even modest military machine vision projects, such as Project Maven’s effort towards autonomous labelling of 38 classes of objects, as these rely on much smaller training databases than their impressive but also fragile civilian counterparts. Despite the virtues of civilian systems – including superior databases and prestigious international competitions running for close to a decade to spur development – non-military systems are prone to not only systematic misrecognition but to systematic misrecognition of fundamentally non-iconic images as objects. It seems, from the account of Project Maven above, that machine vision systems may perform well enough in terms of ‘actionable intelligence’ when and if what is required of such intelligence is crude enough – for example, simple tasks such as following or identifying a car in drone footage from an area without heavy traffic. This militarised version of knowledge seems to bother less with reliability and more with producing possibilities for further ‘action’. Yet the interrogations of civilian systems with far larger and more sophisticated training databases available shows that the form of understanding or intelligence that machine vision systems arrive at is fundamentally unrelated to what we call ‘vision’ in normal human interaction, and grounds this in visual ambiguity being sidestepped in the very fundamental premise of machine vision systems – the construction of training images and labels that enable the system to establish connections. Put in another way, whatever machine matching of images with labels and object categories is, it is not vision. Rather it is statistics, sidestepping the agency of the semiotic act in favour of ‘ground truth’. The introduction of engineering ideas such as ‘ground truth’ into the domain of the always unstable, ambiguous, and culturally specific social conventions that we call vision or seeing is, of course, deeply problematical, as many have long noted. The problems of using anthropomorphic language to describe algorithmic decision-making has been flagged before. Decades ago, McDermott derided, for example, the use of anthropomorphic language to describe algorithmic decision-making as ‘wishful mnemonics’ and warned that they were misleading in suggesting that algorithmic problem solving was analogous to human intelligence. Suchman and Sharkey take up the concept and refer McDermott’s suggestion for de-anthropomorphising algorithmic technologies like machine vision – ‘that we use names such as “G0034” and then see if it is as easy to argue that the program implements “understanding”’.⁹⁹ While this question is amusing, it may turn the contemporary problem in militarised machine vision on its head. It may rather be that when

⁹⁷Allen, ‘Bringing AI to the Fight against ISIS’.

⁹⁸Nguyen, Yosinski, and Clune, ‘Deep neural networks are easily fooled’, p. 2.

⁹⁹N. Sharkey and L. Suchman, ‘Wishful mnemonics and autonomous killing machines’, *Proceedings of the AISB*, 136 (2013), pp. 14–22 (p. 17).

algorithms are already delegated decision-making powers, as they are when they are employed to track vehicles, predict crime,¹⁰⁰ mediate political discussion,¹⁰¹ or function in other areas,¹⁰² anthropomorphic language is employed as part of the ‘discursive work’ of algorithmic governance, deployed to legitimise the decision-making of such systems.¹⁰³ It is striking that while terms like ‘intelligence’, ‘vision’, and ‘recognition’ abound, anthropomorphic description of the functions of algorithmic systems – arresting citizens, making security decisions, or deciding on visa applications – remain rare.

How then does all this match with Haraway’s idea that ‘all eyes, including our own organic ones, are active perceptual systems, building in translations and specific ways of seeing’?¹⁰⁴ Should we avoid terms like ‘seeing’ and ‘vision’ when interrogating what the scatter peacocks mis-recognised by algorithms can tell about technologies entering the battlefield? Perhaps the term ‘sorting’, prominent in surveillance studies, comes closer to something that we can understand as agency that carries a responsibility, yet not mistake that for an act that is intrinsically part of being human (like intelligence, sensing, etc.). At the very least, these controversies show how machine object recognition is not only an active perceptual system but a fundamentally post-human one, developing non-mimetic logics from the training data labelled in a reductionist mimetic designation of visual ground truth aimed at reducing perception to binaries and doing away with the ambiguity that may not be a ‘mistake’ in perception but rather a constitutive part of it. Were we to ask machine vision technology what it does, it would probably agree that vision, with its subjectivity and creativity, is not an appropriate description. The sorting performed relies on very different registers as it can perform on images with no mimetic qualities – as remarked by Armstrong, machines are alone in being able to ‘[v]iew images of oysters, herring, lobster, bread and butter, beer, or wine without experiencing the stirring of appetite.’¹⁰⁵ Vision and seeing were never innocent metaphors, rather as Law¹⁰⁶ and others have long pointed out, these terms have acted as one of the legitimising metaphors to establish science as neutral and detached from that which it works on and transforms. It is for this reason that I believe that the area of phenomena I tell fables about here has to be, at the very least, labelled ‘techno-vision’.

False-colour reversal film and ambiguous detection of deception in Mosse’s ‘Come Out (1966)’

My third entry point into militarised techno-vision is a piece of detritus that was picked up by an artist and then repurposed. A relatively steady stream of critical research has proceeded by using other engagements or controversies to interrogate security technologies and governance shrouded in secrecy, be these engagements scientific,¹⁰⁷ managerial,¹⁰⁸ artistic,¹⁰⁹ or public interest¹¹⁰

¹⁰⁰D. Ensign, S. A. Friedler, S. Neville, C. Scheidegger, and S. Venkatasubramanian, ‘Runaway Feedback Loops in Predictive Policing’, *arXiv:1706.09847 [cs, stat]* (2017).

¹⁰¹Saugmann Andersen, ‘Video, algorithms, and security’.

¹⁰²L. Amoore and V. Piotukh (eds), *Algorithmic life: Calculative Devices in the Age of Big Data* (London: Routledge, Taylor & Francis Group, 2016).

¹⁰³Gillespie, ‘The relevance of algorithms’, p. 180.

¹⁰⁴Haraway, *Simians, Cyborgs, and Women*, p. 190.

¹⁰⁵K. Armstrong, ‘Top 10 things that machines can easily do that humans can’t do’, *For Machine Use Only: Contemplations on Algorithmic Epistemology* (New York: &&& c/o The New Centre for Research and Practice, 2016), pp. 15–17 (p. 15).

¹⁰⁶Law, *After Method*.

¹⁰⁷T. Villumsen Berling, ‘Science and securitization: Objectivation, the authority of the speaker and mobilization of scientific facts’, *Security Dialogue*, 42:4–5 (2011), pp. 385–97.

¹⁰⁸P. Schouten, ‘Security as controversy: Reassembling security at Amsterdam Airport’, *Security Dialogue*, 45:1 (2014), pp. 23–42; M. Tazzioli and W. Walters, ‘The sight of migration: Governmentality, visibility and Europe’s contested borders’, *Global Society*, 30:3 (2016), pp. 445–64.

¹⁰⁹F. Möller, ‘Photographic Interventions in post-9/11 security policy’, *Security Dialogue*, 38:2 (2007), pp. 179–96; Shapiro, *Studies in Trans-Disciplinary Method*.

¹¹⁰W. Walters, ‘Drone strikes, dingpolitik and beyond: Furthering the debate on materiality and security’, *Security Dialogue*, 45:2 (2014), pp. 101–18.

engagements. Similar avenues can be expected to help understand militarised techno-vision and to speak with the technologies used in this militarisation.

Richard Mosse's photo series from Congo, 'Come Out (1966)', 'Infra' and the video installation 'The Enclave' presents an excellent entry point for thinking about militarised visual technologies. In the following, I use the images as devices to pry open military precision optics and lethal machine vision, looking at Mosse's artistic images not so much to see what they show, but rather as what the visual culture theorist W. J. T. Mitchell calls 'picture theory', images that 'develop' not a picture but a theory, much in the same way as a text does. When most powerful, a picture 'offers ... an *episteme*, an entire system of knowledge/power relations'.¹¹¹ I employ Mosse's images to help such theorisation and questioning of the episteme on which military techno-vision relies. It is not certain, nor necessary to my purposes, that this is what Mosse intends with his images. Used as 'picture theory', the images are not to be held responsible for the fables I use them to tell.

The photos and videos in 'Come Out (1966)' from where Figure 4 stems, 'Infra', and 'The Enclave' use Kodak Aerochrome, a discontinued military reconnaissance film with an extra infrared exposure layer. The film's infrared layer enables it to register chlorophyll in live vegetation, and was developed for aerial vegetation surveillance, and deployed to reveal camouflage. According to the producer it is 'an infrared-sensitive, false-color reversal film intended for various aerial photographic applications where infrared discriminations may yield *practical results*',¹¹² or in the terminology used here, provide actionable intelligence. The way in which the Aerochrome film produces these practical results is by colouring plants with living/active chlorophyll – an agent of photosynthesis – pink; while leaving the dead organic or non-organic material used in camouflaging efforts in its original colours.

The film was developed as an enhanced way of military seeing, operative in camouflage detection. It was 'developed by the US military in the 1940s to detect camouflage and to reveal part of the spectrum of light the human eye cannot see'.¹¹³

As explained by the producer,

Color infrared-sensitive films were originally designed for reconnaissance and camouflage detection. In fact, the term 'CD' was once used to denote the camouflage detection role of this film. Color infrared films were sometimes effective when used to photograph objects painted to imitate foliage. Some paints may have infrared reflectance characteristics quite different than those of foliage. In the resulting color infrared transparency, the areas of healthy deciduous foliage will be magenta or red, and the painted objects may be purple or blue. (However, some paints have now been developed with spectral curves closely approximating those of some foliage.) Camouflaged areas are most easily detected by comparing a transparency made on color infrared-sensitive film with a normal color transparency of the same objects on normal film.¹¹⁴

In 'Infra', Mosse uses Aerochrome to take pictures of scenes in eastern Congo, in areas torn by a civil war that for Western spectators offers no easy visual distinction between 'good' and 'evil', the only immediate visible distinctions being between armed and unarmed, uniformed and non-uniformed. This uncertainty serves to distance the Western spectator from the political context of

¹¹¹Mitchell, *Picture Theory*, p. 58.

¹¹²Kodak, 'KODAK AEROCHROME III Infrared Film 1443' (2005), emphasis added.

¹¹³J. Stearns, 'Shocking pink', *The Guardian* (28 May 2011), available at: {<https://www.theguardian.com/artanddesign/2011/may/28/richard-mosse-infrared-photos-congo>}.

¹¹⁴Kodak, 'KODAK AEROCHROME III Infrared Film 1443', available at: {https://www.kodak.com/uploadedFiles/Corporate/Industrial_Materials_Group/ti2562.pdf}.



Figure 4. Photograph from the series 'Come Out (1966)' by Richard Mosse.

the violence and blur our expectations of what we are to see, fostering a somewhat bewildered rather than confirmatory gaze.

In 'Infra', the erratic colouring provided by the Aerochrome film becomes dizzying rather than legible, a destabilising effect of the pink 'magic' of showing the world unknown to us. Mosse has explained that this misfit was what stirred his interest in using the film in Congo, remarking that 'it seemed inappropriate and made me feel slightly uncomfortable'.¹¹⁵ The film makes the images almost surreal, beautiful, perhaps, but also confusing while at the same time showing the militarised production of visual 'facts', the labelling of green in terms of nature and deceptive nature. Mosse's Aerochrome images are opaque and hard to read operationally, their distribution of real and unreal unreliable and confusing rather than readily providing the 'actionable information' touted by military visualisation technology producers. It thus reveals a world of technologically enhanced vision that loyally does the work of distinction that it is constructed to, yet at the same time questions its own applicability. Soldiers are marked as living plant materials, civilian housing stands out marked as deceptive, landscapes full of suspicion meet us and there appears to be at least as many questions after the use of the film as there were before. Where the filmic technology promises to extend our bodily abilities of seeing to leave us with a clearer view of opaque battlefields and 'detect enemy positions in the underbrush',¹¹⁶ Mosse exposes the visual fragility of such a technology that is extremely fragile in both representative and practical ways – it must be kept cold until exposure, and only very few studios can develop it. Observers have called attention to the way in which the resulting photography plays with tropes of Congo and Africa as the 'dark continent',¹¹⁷ and truly it does. But it is also able to 'make us call into question pictures we

¹¹⁵Richard Mosse interview: 'Kodak aerochrome in the Congo', *I Still Shoot Film* (2013).

¹¹⁶Stearns, 'Shocking pink'.

¹¹⁷Lewis Bush, 'Visions of War: Richard Mosse and The Enclave' (2013).

thought we understood’,¹¹⁸ and, I think, asks us to question their apparatus of production, the militarised technologies we are being asked to believe in but which we are not given access to scrutinise. The lethal ways of seeing enacted in militarised visual sensing systems depend on our trust in the visual technologies as ways of knowing the social or natural world without too much ambiguity. Such trust, then, forms the basis for the enactment of a ‘sovereign gaze’ that can support security decisions.¹¹⁹ With Mosse, we can scrutinise a warzone as the militarised visual technologies of the past would do, and appreciate that not only is technological vision transformative, it is collaborative: it contains an inescapable and hidden layer of uncertainty, confusion, and mess that is being erased in the conversion of seeing – an always subjective, deeply personal, and unstable biological-cultural faculty¹²⁰ – into a militarised form of knowing. This conversion is something that the technology cannot do, but where its black-boxed agency enables its operators to step in and gloss over ambiguity with technological certainty.

The enactment of a division between the real and the unreal can, I argue, be made visible in the way in which we can leverage Mosse’s use of the Aerochrome film to ask questions about the Aerochrome technology and militarised optical technologies more broadly. As citizens in whose name technologies such as Aerochrome are deployed but who are *not* allowed to be spectators to their deployment, we are told that this is a technological way of seeing the difference between nature and faked nature, revealing deception. But in the warzone of Congo, where kids are represented as playing next to corpses, what we see is not a fresh perception of a world made actionable in terms of real and fake by militarised optic technology, but a hardly intelligible, beautiful mess stemming from the technology rather than alleviated by it. Mosse, in a short interview film, states this as his ambition at least to a degree. The beauty of the images of the Congo conflict, he explains, ‘creates an ethical problem in the viewer’s mind so then they’re like confused and angry and disorientated and this is great because you’ve got them to actually think about the act of perception and how this imagery is produced and consumed’.¹²¹ While this may be the case for a photographer, the beautiful mess can also be used to think about the chain of security knowledge/decisions such images enter into. This points to the agency of the technology (as making a beautiful mess) but also to how technology is not lethal vision in itself: it becomes painfully clear that what comes after the technological translation of the world is a highly productive enactment of the mess as security knowledge – actionable intelligence – stripped of the ambiguity that abounds after techno-visual intervention. In the human-machine reconfiguration enabled by this technology, we are not beyond the need for interpretation and decision-making, rather we are in the middle of it.

Conclusion: Ambiguity and actionable intelligence in militarised techno-vision

In this article, I have tried to compose the composite object of military techno-vision from public documents, doing so without reducing the ambiguity and heterogeneity of its constituent components, without aiming or claiming to exhaust what these constituent components may be, and without claiming that the story could not have been told otherwise or that the composition presented here is infallible. The need for such an eclectic composition from socio-technological fables, rather than a neat, coherent, and rigorous analytical framework, stems, to me at least, from the performative power of the black boxes in which military techno-vision is cloaked. Rather than just an unfortunate side effect of unavoidable security concerns, such black-boxing is itself a productive force which directs analysis to either engage in implicit bargaining over and gratitude for access, or to the use of available material that is often promotional. In the composition of military

¹¹⁸Stearns, ‘Shocking pink’.

¹¹⁹L. Amoore, ‘Vigilant visualities: the watchful politics of the war on terror’, *Security Dialogue*, 38:2 (2007), pp. 215–32.

¹²⁰D. Lyon, *The Electronic Eye: The Rise of Surveillance Society* (Minneapolis: University of Minnesota Press, 1994).

¹²¹Frieze, *Richard Mosse: The Impossible Image* (2013), 6:07.

techno-vision, layers of secrecy and opacity works in tandem with promotional discourses of artificial intelligence, more-than-human vision and (surgical) precision, legal and political discourses of less-than-civilian subject positions such as military-aged-ones or people 'bringing their kids to the battlefield',¹²² and with regimes of visibility in commercial media and visual culture. This enables the composition and legitimisation of military techno-vision as technologically and ethically superior production of 'easily detected' facts yielding 'practical results', 'ground truth' in an unambiguous domain of statistical techno-vision, and enables detritus from the actual operations of militarised visual technologies to be dismissible as occluded by a 'fog of war'. Picking up such detritus, telling its stories, and asking about the work the technology it stems from is doing for us, asked to do for us, and about the configurations of humans and machines around it, I have shown, enables a composition of military techno-vision as complex configurations of humans and technologies in which decisions and actionable intelligence are always shrouded in ambiguity, and where ethics cannot be delegated to or vested in technologies. The desire to obscure ambiguity and render technological vision as actionable intelligence is visible in the interpretative anchoring applied to PR-videos as well as leaked video, in the brushing off of inconvenient leaked images as requiring professional vision to be interpreted, in notions of ground truth, and in the visualisation of a surreal mess produced in the application of military techno-vision to the battlefield in Congo.

In different ways, the heterogenous pieces of detritus examined here tells us that ambiguity remains constitutive of both images and perception, and is merely glossed over in terms such as actionable intelligence that are mobilised to sustain security decisions and to render these potentially lethal decisions rational and reduce the ethical gravity they carry. This 'hygienic'¹²³ performance of the elimination of ambiguity is at the core of an epistemic politics of military techno-vision, manifested in different ways in the heterogeneous components that make up the area.

Acknowledgements. With the usual caveats regarding the final product, I wish to thank the Special Issue editors, journal editor, anonymous reviewers, insightful participants at the 'Compositions' workshop in San Francisco, and at the TASER seminar in Tampere in 2018, Mareile Kaufmann and other encouraging critic peers at NordSTeva 2017, as well as a perceptive audience and fellow panellists at the 2018 EISA and 2019 ISA panels where this work was presented.

Rune Saugmann is an Academy of Finland postdoctoral researcher in the International Politics programme at University of Tampere, and Docent in Media and Communications Studies at the University of Helsinki. His interdisciplinary research on the visual mediation of security has investigated how videos participate in constituting security, and is increasingly turning to how digital images are used for security purposes in computer vision systems. His work has appeared in *Security Dialogue*; *European Journal of International Relations*; *Journalism Practice*; *European Journal of Communication*; *Global Discourse*; *JOMEC*; the British Computing Society's Electronic Workshops in Computing series, as well as numerous edited collections in IR and media studies. Saugmann is a driving force in developing visual ways of thinking about International Relations and is the editor (with Juha Vuori) of *Visual Security Studies* (2018), the first volume dedicated to the visual study of security. His research is available at saugmann.tumblr.com

¹²²A comment made prominent in WikiLeaks, 'Collateral Murder' (short version).

¹²³Law, *After Method*, p. 10.