CHAPTER 4

An Alternative Rationalisation of Creative AI by De-Familiarising Creativity: Towards an Intelligibility of Its Own Terms

Jenna Ng

'There, look!' we could say. 'Look at this art! How dare you claim these children are anything less than fully human?'

– Kazuo Ishiguro (2005, 238)

Introduction

This chapter formulates an alternative understanding of creative Artificial Intelligence (AI) by examining how the computational terms of AI may be rationalised in a framework *intelligible* to humans. The level of algorithmic processing today presents two tensions which hinder a full comprehension of creative AI. The first is the still formidable lack of transparency of AI's workings, as noted by many scholars *à la* the algorithmic 'black box' (Pasquale 2015; Diakopoulos 2016; Brill 2015; Ananny and Crawford 2018). The second is the increasing lack of human intervention in the algorithm's processing not only through the seemingly unfathomable operation of its 'black box', but also through algorithms learning from other algorithms, such as by way of a Generative Adversarial Network (GAN). The result is to consider anew how computers may be

How to cite this book chapter:
Ng, J. 2021. An Alternative Rationalisation of Creative AI by De-Familiarising Creativity: Towards an Intelligibility of Its Own Terms. In: Verdegem, P. (ed.) *AI for Everyone? Critical Perspectives.* Pp. 49–66. London: University of Westminster Press. DOI: https://doi.org/10.16997/book55.d. License: CC-BY-NC-ND 4.0

considered to be autonomous creators – to be genuinely creative in and of itself, or creators 'in their own right' (Veale and Cardoso 2019, 2). Or, as Hofstadter (2000) writes: 'It [the mechanical substrate of creativity] may not constitute creativity, but *when programs cease to be transparent to their creators*, then the approach to creativity has begun' (669; emphasis added).

Recent innovations in creative AI bear out both tensions, where the algorithm generates creative decisions, say on note, word or paint placement, out of its *own* processing of the dataset it receives, and in ways not entirely understandable to humans. This level of processing may be contrasted with how computers had previously produced creative work, or in what is known as automated creativity. As early as the 1950s, computers have produced creative outputs, such as music, by running codes of basic material and stylistic parameters which enabled the generating of 'raw materials'. These musical 'materials' were then modified and assembled by human composers into recognisable pieces of music (Alpern 1995). In these cases, the computer was *specifically programmed to produce the creative output*, following instructions on where and how to place notes or dabs of paint, even if those instructions may be rules of randomness.[1] The research field of Computational Creativity recognises such programming as 'pastiche', where the computer's creativity is a 'mere appearance', and only 'due to some specifiable slice of the programmer's own creativity having been imprinted onto the algorithmic workings of the system'. (Veale and Cardoso 2019, 4).

Conversely, current creative AI operates as neural networks which discern specific patterns out of processing large datasets of relevant outputs, thus 'learning' across complex nodal networks of 'reward' and 'punishment' the placements of note, paint and words for producing the creative output in keeping with the patterns they have 'learned'. A couple of examples to illustrate this: in 2016, the Project Magenta team at Google unveiled a 90-second melody produced by a computer to which they fed 'some 4,500 popular tunes' and 'seeded' with four musical notes. By processing the large database of tunes, the network 'composed' its melody by discerning patterns of musical rules and constraints *in ways never specifically programmed into it*. The algorithm may also learn from other algorithms, such as the algorithmically-produced portrait of Edmond Belamy in 2018. In this case, a discriminator algorithm was first trained on a database of 15,000 portraits (painted by human artists) that had been uploaded to it (by human computer scientists). The discriminator algorithm was then used to 'train' a separate generator algorithm which 'learnt' paint placements and so on based on 'reward' and 'punishment' feedback from the discriminator algorithm, both doing so through processing enormous amounts of data. The Edmond Belamy painting later made history as 'the first portrait generated by an algorithm to come up for auction', and eventually sold by Christie's for a not insubstantial sum of US$432,500 (Cohn 2018). The key issue in these cases is that the algorithms have not been specifically programmed to place notes and paint; instead, they 'learnt' to do so through processing enormous amounts of data and being given signals on what placements were 'correct' or 'wrong'. On that feedback, they then generated their respective outputs.

While not quite the spectre of a Terminator machine out to annihilate the human race, creative AI on these terms is disturbing in how our lack of understanding of its creativity and creative process reinterrogates our notions of humanness, where creativity has always been its indisputable hallmark (Zausner 2007). It is 'part of what makes us human' (Sawyer 2006, 3), and affirms our humanity (Csikszentmihalyi 1990); it colours the domains in which humans work, think, play, produce and perform (Kaufman and Baer 2005). Per the opening quotation of the Introduction, the clones in the speculative society of Kazuo Ishiguro's (2005) acclaimed novel, *Never Let Me Go*, made art as a concerted attempt to evidence their humanity. As their teacher explained to them: 'we thought it would reveal your souls. Or to put it more finely … to prove you had souls at all' (Ishiguro 2005, 238). The clones' creative work were sought to demonstrate humanity, for 'the creativity code', as Marcus du Sautoy (2019) puts it, 'is a code that we believe depends on being human' (2–3). How, then, may we understand so critical a touchstone of humanness in AI when creativity is seemingly manifest on such opaque and unintelligible terms?

This chapter thus proposes *a framework of de-familiarisation* for the paradoxical task of rendering the computer's creativity, seemingly so entrenched on its own terms of computational data and processing, intelligible in human terms. Its aim is to propose an approach with which to rationalise the processes of the computational algorithm, if anything to render the clarity of the imbrications between the human and the computational that colour so much of our algorithmic world today. First, as a brief literature review, I present a few salient tenets of existing rationalisations of AI. In particular, I critique how their approaches, by and large, extract comparative analyses between human functions and computational processes. To formulate an alternative approach, I then draw from rationalisations of media out of media theory, specifically theorisations of the marionette by Heinrich von Kleist (1810) and of the camera via Russian filmmaker Dziga Vertov's (1923) writings on cinema, to present a methodology of *de-familiarisation* as an approach to rationalising technology *on its own terms*. In the third section, I apply that perspective to re-think creative AI via the case study of AlphaGo, an algorithm programmed by Google DeepMind to play the game of Go and which made AI history in 2015 by becoming the first computer programme ever to beat a human professional player at the game. While AlphaGo does not produce artistic work *per se*, it serves as an apt case study as its moves were deemed to be of exceptional novelty – indeed, described as 'creative genius' (Sautoy 2019, 34) – and in various ways considered to have re-defined the frontier of AI. The last section will conclude. The uniqueness of this argument thus lies in how it aims to shift the conversation from an us-and-them framework, where computing is often conceived on the singular oppositional dimension of humans versus machines (such as comparing computers directly against humans). This alternative approach to understanding algorithms thus suggests a different dimension to that understanding – not one made on human terms, but as a paradoxically impossible approach of the algorithmic being humanly intelligible on its own terms.

Current Rationalisations of Creative AI

Current consideration of creative AI lies in extensive scholarship, not least because much of it sits within a vastly wider enquiry: can computers be human? In the face of this question, current discourse inevitably turns to a comparative methodology, whereby the computer's processes are compared against multiple manifestations of human cognitive function, including creativity (Dreyfus 1972; Dreyfus 1979; Bailey 1996; Moravec 1998; Boden 2004). Various conclusions are then reached by matching one against the other, and working out how each measures up.

In other words, the rationalisation of AI is laid out in *comparative* terms, so that AI becomes intelligible only as *against human capacities*, or against what AI can or cannot do as compared to humans. The multiplicities which reflect this rationalisation across philosophy, computer science, cognitive psychology, cybernetics, neuroscience and myriad other disciplines are myriad and intricate, and far beyond the scope of this chapter to cover comprehensively. A few highlights will hopefully suffice to demonstrate its contours. We might, for instance, think about Vannevar Bush's famous imagining of a memory machine he named the 'memex' (Bush 1945), influential to the present day as a basis for the World Wide Web (Davies 2011). Notably, Bush presented the memex as a technology directly against human memory, specifically referencing the former's mechanised processes of speedy and flexible consultation and storage against the latter's corresponding weaknesses, leading to impermanence (forgetting) and lack of clarity (confusion). Conversely, Bush also noted the strength and speed of association of human memory, concluding that 'man cannot hope fully to duplicate this mental process [of association] artificially, but he certainly ought to be able to learn from it' (Bush 1945, n.p.). Both points illustrate Bush's rationalisation of technology as a counterbalance to human capability, whereby one variously contrasts against, supplements, and demonstrates differences against the other. The technology is thus made intelligible as *against the human*, specifically in terms of what it can augment and surpass, and what it cannot.

As AI – itself the field of computers which simulate human cognitive capacities – increases in operative sophistication to resemble human intelligence, this contrast becomes ever more explicit. The Turing test (Turing 1950), even more famous than Bush's memex machine, reconciled computer cognition in terms of whether it was distinguishable – or not – from human behaviour.

Indeed, John Searle, among many other computer scientists, distinguished between 'strong AI' and 'weak AI' in his now classic 1980 paper, 'Minds, Brains, and Programs' (Searle 1980), later developed into his book, *Minds, Brains and Science* (Searle 1984), precisely on such rationalisations of the computer against characteristics of human cogitation. In his paper, Searle argued that AI, in its state of development then, could only be 'weak', whereby 'the principal value of the computer in the study of the mind is that it gives us a very powerful tool'. Conversely, it was not 'strong', whereby 'the appropriately programmed computer really *is* a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states … the programs are themselves the explanations' (417; emphasis in original). Searle clinched his argument against 'strong AI' by arguing its lack of free will and other mental states which, according to Searle, characterise human cognition, thus demonstrating the limitations to 'computer simulations of human cognitive capacities' (417). Again, these arguments rationalise AI against the human in a comparative mode. They render AI intelligible by referencing its technological Otherness against constructed definitions of natural human responses.

As a counter-stroke, we might also think about the extensive work in computer and cognitive sciences which rationalise *the human being* in computational terms. However, this shifted understanding of the human as a computer only expands the commensurability between AI and humans, this time not by *difference* (humans against computers), but by *equivalence* – humans are computers. In turn, this intelligibility of AI via counterpointing the human – in terms of underscoring AI's logical and mechanised processes as against the biological and the organic – expands to not only the rationalisation of the human, but the world itself. This, then, is the core of *computationalism*, defined by Golumbia (2009) in its 'received', or 'classical', form, as

> … the view that not just human minds are computers but that *mind itself* must be a computer – that *our notion of intellect is, at bottom, identical with abstract computation*, and that in discovering the principles of algorithmic computation via the Turing Machine human beings have, in fact, discovered the essence not just of human thought in practice but all thought in principle (emphasis added). (7)

The idea has flexed and flagged in multiple ways and forms, but its central concept remains the conceptualisation, understanding and identification of human cognition and mind in computational terms. In this respect, we can also think about, for instance, Allen Newell and Herbert Simon's work across the decades from the 1950s which specifically argued for the model of all human reasoning to be representable as symbolic 'information processing systems' (Newell and Simon 1972). Giants in their respective fields, both awardees of the Turing Award and Simon as well a Nobel laureate, their thinking converged with others at the Dartmouth summer conference of 1956,[2] whose specific mission, notably, is 'to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it' (McCarthy et al. 1955, n.p.). As the field (and the term) of 'Artificial Intelligence' emerged out of the Dartmouth conference, its ideas folded into another new field developed in the 1970s – namely, cognitive science, which studies human thinking, learning and perception as coloured by cybernetics, neuroscience, linguistics and psychology, but also dominated by AI, mathematics and computation (Gardner 1985). The field of cybernetics, first emerging out of the Macy Conferences on Cybernetics from 1943 to 1954, also forged new paradigms out of information theory, neural functioning and computer processing, among others, to become 'a new way of looking at human beings. Henceforth, humans were to be seen primarily as information-processing entities who are *essentially* similar to intelligent machines' (emphasis in original; Hayles 1999, 7).

This approach of rationalising the human in computational terms infuses much of current thinking about creative AI (Boden 1996; Boden 2004; Miller 2019; Sautoy 2019; Kaufman and Baer 2005). The definitional knottiness of the term 'creative' aside – over 60 definitions of 'creativity' appear in psychological literature alone (Boden 1996, 268) – the broad rationalisation of creative AI continues along comparisons against human creativity as couched *in computational terms*. Hence, for instance, Douglas Hofstadter suggests the 'mechanisation of creativity': while 'creativity is the essence of that which is not mechanical', '[y]et *every creative act is mechanical* – it has its explanation no less than a case of the hiccups does' (emphasis added; Hofstadter 2000, 669). Similarly, Herbert Simon, as already seen, justifies human cognition as informational processing systems, and thus posits that 'creativity involves nothing more than normal problem-solving processes' (as quoted in Csikszentmihalyi 1988, 19). More recently, Miller (1992; 2000; 2019) rationalises human creativity as 'a model for network thinking' in terms of 'many lines of thought taking place at once in parallel, coming together from time to time to enrich each other' (Miller 2019, 36). The model thus demonstrates how 'new ideas do not just pop up out of nowhere, even though they may seem to', thereby visualising human creativity as a mappable process that is also reproducible on a computer (Miller 2019, 29). Having said all that, the idea of mechanised creativity stretches back to the ancient Greeks, where Burkholder, Grout and Palisca (2014), for instance, argues that Pythagoras and his followers held that 'numbers were the key to the universe', and thus thought music as 'inseparable from numbers' (13). Creating music, then, was really the theoretical application of numbers and various mathematical properties in logical and calculable steps, not unlike in algorithmic fashion.

Across these formidable rivers of thought and expansive arguments, if sketched in generous outlines and overlooking many more, we may thus identify how rationalisation of human creativity along mechanical computational processes dovetails into the persistent thinking of AI as made intelligible against the human. As such, the intelligibility of AI continues to be tracked against shifting interpretations of human capability and, in that respect, facile notions of what constitutes humanness. The arguments which directly oppose creativity as defined in terms of logic and mechanisation just as easily deploy a notion of creativity that appeals to other touchstones of humanness deemed still unachievable by the computer, such as consciousness, self-understanding and awareness: a truly creative computer, after all, 'cannot be a dumb savant that naively flings outputs at an audience' (Veale and Cardoso, 2019, 4). Or that creativity entails unique human experiences, such as 'the need for experience and suffering' (Miller 2019, 16) or self-actualisation, where creativity is 'about humans asserting that they are not machines', and 'to expose what it means to be a conscious, emotional human being' (Sautoy 2019, 283). Hofstadter (2000) refers to 'the depth of the human spirit' for a meaningful notion of creativity:

> A 'program' which could produce music as they [Chopin or Bach] did would have to wander around the world on its own, fighting its way through the maze of life and feeling every moment of it. It would have to understand the joy and loneliness of a chilly night wind, the longing for a cherished hand, the inaccessibility of a distant town, the heartbreak and regeneration after a human death. It would have to have known resignation and worldweariness, grief and despair, determination and victory, piety and awe. (672–673)

The point here is not to argue for any particular definition of or position about the computer's creativity. Rather, it is to underscore the various lenses deployed through which AI creativity is rendered intelligible, namely, *in relation to the human* in terms of comparison, contrast and analogy. In some ways, this is wholly intuitive – as mentioned, technology forms a counter-distinction to humanness; it mirrors the age-old binary of artifice against nature. It thus makes sense to employ humans as the referential framework in understanding the technological Other. Yet, the approach is also flawed. Understanding AI on these terms becomes subject to changing constructions and perspectives of humanness, so that it relies on a precarious balance of what AI *is* against what it is *not*. Conditional on being defined in relational terms, it fails to have its own definitional footing. An understanding of AI on these foundations cannot be a thorough one.

Moreover, this approach of comparison confines our understanding of AI to being *within the intelligibility of human terms*, rather than made intelligible *on its own terms* as a logical, mechanical and computational entity. This is important because at the heart of an intelligibility on human terms is an incommensurability that is never truly addressed: computers are simply not humans. A comparison that renders one intelligible on the terms of another will always lose something in the translation. The next section, drawing on alternative perspectives from theorisations of media, will suggest a different approach.

De-familiarising Creativity

As a starting point, much of media technology is similarly rationalised as comparisons against human capabilities, as this broad scattering of examples will hopefully suffice: Nicholas Carr (2010), for instance, in resonant echoes of Marshall McLuhan ([1964]; 2013) rationalises technology as an expansion of 'our power and control over our circumstances' (44), such as the map and the clock which 'extend and support' the 'mental powers' of humans in formulating, producing and sharing knowledge. Jonathan Safran Foer (2016) writes of communication technologies as 'substitutes' for real-time face-to-face human interaction: 'We couldn't always see one another face to face, so the telephone made it possible to keep in touch at a distance. One is not always home, so the answering machine made a message possible without the person being near their phone.' (n.p.). Edward Branigan (2006) suggests anthropomorphism as an 'analytic category' to measure 'the degree to which a camera is being used to simulate some feature of human embodiment', whose analysis then relates the qualities of the camera to 'a typical way of human viewing, or moving (or thinking and feeling), and to what degree' (37). William Brown (2009) argues the pole converse in his thinking about 'posthumanist cinema', demonstrating how the digital 'posthuman camera' *omits* human embodiment entirely in its humanly impossible shots.

This comparative approach is also an old rationalisation. In his 1880 essay, Jean-Marie Guyau analogizes the human brain to the phonograph, drawing connections between recorded sound as grooves of vibrations engraved onto the phonograph's metal plate and the 'invisible lines [that] are incessantly carved into the brain cells, which provide a channel for nerve streams' (31). Guyau further analogizes the speed and strength of vibrations of our brain cells in terms of images conjured by our minds to the speed of the phonograph's vibrations and the tones of its sounds. In 1950, George Wald, a professor of biology at Harvard, noted resemblances between the camera and the human eye – 'of all the instruments made by man, none resembles a part of his body more than a camera does the eye' (32). Wald (1950) further detailed similarities between human vision and photography: 'the more we have come to know about the mechanism of vision, the more pointed and fruitful has become its comparison with photography' (32). He described how the chemical changes of exposed photographic film, particularly 'dark reaction' of the 'latent image' in the darkroom, mirrors processes of vision in the eye's exposure of rhodopsin[3] to light (40). Along these broad contours, the rationalisation of media technology thus echoes that of AI – as matched against human capabilities, reflecting similarities and differences; as situated *in human terms*.

However, large swathes of critical theory have grown to critique the human as a referential framework. Specifically, this work decentres the human by shifting the critical lens to those that are not human, such as 'understood variously in terms of animals, affectivity, bodies, organic and geophysical systems, materiality, or technologies' (Grusin 2015, vii). Understanding the Other and accommodating *on their terms* their complex involvement in the consideration of our world thus stands as a long-established enquiry through the humanities. This work spans across multiple areas, such as the social imaginaries of inanimate objects (Appadurai 2014); the posthuman (Braidotti and Hlavajova 2018); the nonhuman (Grusin 2015); post-anthropocentrism (Parikka 2015); the intelligibility of cinema as both subject and object of vision (Sobchack 1992) – to cite again just a few sprinklings as illustration. There are many others.

There is, of course, an inherent contradiction to this approach, which is of intelligibility having to be made intelligible in alien terms, or in terms of an Other-ness that, by definition, we do not and are unable to possess. How might we render something intelligible in its own terms if it is, by definition, outside the intelligibility of our own terms? How do we accommodate our understanding around something that we are not, let alone fathom its terms? It is certainly a valid conceptual difficulty. The key is to understand the enquiry not as a literal one which seeks literal answers. Rather, it is one which involves speculation and imagination in envisioning the perspective of the Other as part of its methodology. It requires the acceptance of the philosophy of things being in themselves, beyond and independent of our experience (Moffat 2019). It entails being open to indeterminacy and contingency, and of acknowledging the nature of things as based on, while not pure fantasy, nonetheless an inexact science.

The approach proposed here, then, for shifting one's critical perspective in relation to an intelligibility of the technological Other is to draw on media theory's alternative rationalisation of technology, namely, through *de-familiarisation* – to disturb or disorder the terms in which we think of an entity so as to re-learn it on different terms, specifically those of its own. I underscore two examples, each of a different media through different rationalisations, to more fully illustrate this approach. The first is Kleist's 1810 essay on marionettes, in which he recounts a conversation with his friend, 'Herr C.', who expressed admiration for the gracefulness of the puppets. This position is counterintuitive: puppets, controlled by their puppet-master, are mechanical and lifeless; it is senseless to consider puppets as graceful as, if not more graceful than, human dancers. The key in 'Herr C.'s' reasoning lies in how he *re-reads* the puppet's mechanical movements not as cold actions with no consciousness or with a surrendered volition, but as precisely the nonconscious and mechanical movements *that only puppets are capable of*, through which beauty and grace *re-emerges*. The puppet's artificial properties are thus read *on their own terms* – not against the human dancer's consciousness of its movements which renders their artistry and beauty. Rather, 'Herr C.' *de-familiarises* what and how we think about grace, and relearns it in the non- or unconsciousness of the puppet's mechanical operations. We thus come to a different understanding of the puppet by emerging on the other side of its paradox (i.e., of grace from the controlled and the automatic) to arrive at an alternative intelligibility of it and its movements. *We understand the puppet on its own terms*.

The second example is drawn from Russian filmmaker Dziga Vertov's theorisation of cinema. Articulated in the 1920s through various pamphlets, articles, manifestos and public addresses, and fresh from radical societal change in the wake of the Russian Revolution, Vertov sought from cinema and its camera the newness of humanity and society. He read the camera through a new intelligibility – not against the human or in comparison to the human eye, but on its own terms as what he calls a 'kino-eye' in *how it sees a different world*: 'I am a mechanical eye. I, a machine, show you the world *as only I can see it*' (emphasis added; Vertov 1923, 17). Of course, the human hand and eye still control the film camera; it has not literally come alive. But the argument here is not a literal one. Rather, it is a theoretical shift involving imagination and inventiveness to acknowledge the new-ness of the camera's vision and its alien visuality. Like Kleist with the puppet, Vertov came to understand the camera *on its own terms* as he sought a visualisation of the new society birthed from revolution out of its camera eye: its alien-ness as an un-human consciousness is precisely why the kinoeye is capable of ideology to present the real in a way the human eye cannot. He could thus acknowledge the camera's *different-ness* – as he writes, 'it is the realization by kinochestvo[4] of that which cannot be realized in life' (Vertov 1922, 9). Hence, through re-reading the camera on its own terms, Vertov *de-familiarised* the world around us, presenting it anew in a radical language borne out of the camera's foreign intelligibility, and shifting images out of the referential framework of human seeing.

The point here is not to agree or disagree with Vertov or Kleist in their respective readings of media. It is to illustrate how a frame of reference in understanding can shift with a different rationalisation, and in so doing recognise a different intelligibility. The task, then, is to apply this approach to understanding creative AI on their own terms, to which the next section will turn via the case study of AlphaGo.

Rationalising the Creativity of AlphaGo

AlphaGo, an algorithm programmed to play the game of Go, achieved global fame in March 2016 by defeating Lee Sedol, a highly ranked South Korean professional Go player and 18-time world champion, 4 games to 1 in a 5-game tournament. Its victory sent ripples through the AI community and the wider public because 'teaching computers to master Go has long been considered a holy grail for artificial intelligence scientists' (Yan 2017, n.p.). The difficulty of this 'holy grail' lies in the game's high level of abstraction. Played by two players each placing, in turns, stones of their respective representative colour (black or white) on intersection points between horizontal and vertical lines marked on a board, there are essentially just two rules of play: one on how to 'capture' an intersection; the other on how that intersection is considered 'occupied'. The goal is simple: to have, at the end of the game when all intersections have been 'occupied', stones of your colour 'occupy' more intersections (or territory) than your opponent's.

Like all good Zen koans, its minimalism is also its complexity. Compared to Go, chess, as a fellow strategy game, is clearer in various ways: fewer moves can be made to start a chess game, and thus relatively fewer possibilities branch out from each opening move. Pieces also have set values (the pawn, for instance, has the lowest; the queen the highest) which makes an unfinished chess position relatively easy to calculate and analyse as to which player is winning based on how many and which pieces are left, plus any positional advantages. In comparison, because all there is to Go are stones on line intersections, the result is many more possible board configurations, each one lending themselves to even more possible positions if calculated further down the line. As a result, there is quantitatively *and* qualitatively more ambiguity in Go, with ensuing greater difficulty in analysing who is winning from an unfinished game position. Hence the significance of AlphaGo's victory: due to its multiple positional possibilities – as has been oft-quoted, there are 'more possible configurations of the board than there are atoms in the universe' (Yan 2017, n.p.) – until AlphaGo's triumph, the game was considered unconquerable by computers over human players simply because its level of complexity needed it to be played with human abilities of intuition and grasping of visual structure, rather than the computer's powers of searching and calculation of variations.

Pertinently for our purposes here, AlphaGo was vaunted for not only its tournament victory, but also the creativity of its moves. One move – Move 37 of Game 2 – in particular was so wholly unexpected that commentators described it, if a tad gushingly, as 'a truly creative act' (Sautoy 2019, 37); or 'one of the most creative [moves] in Go history' (Tegmark 2017, 89). AlphaGo's Move 37 was to place a stone on an intersection on the board's fifth line, a move very seldom played at that stage of the game because it was considered too 'high' on the board, giving the opponent room to play on the fourth line down and thereby gain too much solid territory. The media quickly attributed the move to the algorithm's *own* creativity, lauding it, with embarrassing hyperbole, as 'the move no human could understand' (Metz 2016, n.p.). Or, as widely quoted from Fan Hui, the European Go champion who was the first professional Go player to play and lose against AlphaGo: 'it's not a human move. I've never seen a human play this move.' (as quoted in Metz 2016, n.p.).[5] What validated the unusualness of the move – and thus rendering it 'creative' rather than 'insane' or 'nonsensical' – was that, some fifty moves later, that fifth-line stone became an unexpected linchpin to a battle for territory which started in a different part of the board. In due course, the battle joined up with the Move 37 stone, giving AlphaGo the advantage and eventually the win.

As expected, the rationalisation of AlphaGo's Move 37 lay in the conventional framework of human terms via comparison and contrast, this time by placing the algorithm in its own intelligibility as one outside human sense. Yet that does not achieve much for understanding AI in its conceptual sense – it merely blankets the algorithm with mystique of the technological and a cryptic referencing of its Other-ness on the basis of some kind of mysterious agency. For instance, much was made of AlphaGo's independent learning from its neural network to generate its moves. Like the painting of Edmond Belamy whose generator network 'learnt' paint placement from the discriminator network, AlphaGo, as the generator network, 'learnt' the best moves in Go by playing multiple games against another neural network. While the discriminator network would have been 'trained' to play Go by being fed (by human computer scientists) millions of games (played by human players) as downloaded from the internet, it is the processing of the millions upon millions of games between AlphaGo and its discriminator network that makes up AlphaGo's main 'training', namely, the calibration of the values and weightage for its nodes across its various networks which ultimately generates AlphaGo's moves.

The implication, then, is that the algorithm's creativity in coming up with unusual moves is its own, generated on its own steam and out of its own learning, an idea its Google DeepMind creators were keen to perpetuate. For instance, in a video interview with CNBC, Demis Hassabis, CEO of Google DeepMind, implies the same generative creativity, explaining how algorithms such as AlphaGo 'learn from scratch, learn from *their own mistakes* … they learn from themselves, directly from data or from experience, *rather than being told what to do by human programmers*' (emphasis added; as linked in Yan 2017, n.p.). DeepMind have since developed AlphaGo's successive algorithms, AlphaGo Master and AlphaGo Zero, along similar lines, namely, to 'learn' Go rules without any human guidance, but simply through processing millions and millions of games against another neural network, whose 'reward' and 'punishment' outcomes would thus train the algorithm on the rules and optimisation of gameplay (Silver et al. 2018).

The case here, then, is to re-think AlphaGo's creativity *on its own terms*, as with Vertov and Kleist's up-ending of grace and visuality in relation to marionettes and cinema. Here, we re-orientate the thinking of AlphaGo's creativity from its comparisons against moves by human players to *de-familiarise* its creativity so as to stand on its own terms. Move 37 was not generated on non- or unhuman terms as an alien stroke of creativity; it was *calculated* out of multitudinous values and possibilities arising from that particular position, and then chosen as the one which gave it the highest chance of ending up with more territory and thus a win.

But 'creativity' here, as framed on the algorithm's own terms, is not only its multi-layered[6] levels of calculation of the multitudes of moves from the multitudes of board positions to the multitudes of possible future board positions. Such level and extent of calculation resonate with the earlier discussion on creativity as an account of mechanised thought and logic applicable to human cognitive systems. There is a further nuance here: the algorithm's sense of creativity does not just lie in this manifold expansion of logical thinking (which might indeed be traced back to human interventions in training the discriminator network); it is also about the *speed* of its calculation through the multitudes of board position data. Speed, then, is really about space, or the demolition thereof, *à la* Paul Virilio (1991) who calls speed 'a primal dimension that defies all temporal and physical measurements' (18), and which directly results in 'the crisis of the whole', whereby the substantive, homogeneous and continuous gives way to the fractional, heterogeneous and discontinuous. Without veering too much into Virilio's ideas on speed which include the city, urban architecture and media, we can draw this line of thought on speed and space back to the *de-familiarisation* of creativity, whereby the terms of the algorithm are thus about neither its anthropomorphised independent learning nor even its mechanisation of logic and thought. Rather, they are about the algorithm's *speed* as the fracturing of the space-time of thought and as mirrored by the multiple splitting of its tree searches that is key to its algorithmic operation, and further constantly mapped with the algorithm's training from its datasets. Thus de-familiarised, we can shift our conceptualisation of creativity from cognitive processes in human terms to a different framework of space and spatial dimension across which the algorithm traverses with speed. The more tightly controlled the space is with unambiguous rules and outcomes, the more it is suited for discontinuous and fractional spaces, and the more the algorithm will thrive. The rationalisation of its 'creativity' in generating brilliant moves with large probabilities of achieving game-winning positions is thus not based on the depth of its logical thinking and learning as per the terms of rationalised human creativity. Instead, based on its computational calculatory processes, we can read it as a more de-familiarised conceptualisation of *space*, and begin the course of understanding it on its own terms.

Conclusion

> *The man-machine assemblage varies from case to case, but always with the intention of posing the question of the future.*
>
> – Gilles Deleuze (1985, 263)

Intelligent AI – more specifically, runaway intelligent AI which not only surpasses humans in general intelligence but whose capabilities are no longer under human control – has been identified as an existential risk, capable of wiping out the human species (The Economist 2020b; Bostrom 2014). Stephen Hawking has pronounced to the BBC that 'the development of full artificial intelligence could spell the end of the human race' (Cellan-Jones 2014, n.p.). AI represents profound fears – culminating in our extinction – but also profound hopes in bettering life for humanity and life on Earth.

For these reasons – to ward off our fears and harness AI for betterment – the need to continually push for deeper understanding of AI is also correspondingly clear. The current rationalisation of AI persists in human terms, as is evident from even the most recent musings on the limitations of AI (The Economist 2020a): comparisons are consistently made with human learning and cognition, such as 'embodied cognition', or references to the 'irredeemably complex' nature of human minds (n.p.). In filling the gap of understanding why AI is still rubbish at doing elementary tasks that humans accomplish without much thinking, such as recognising a stop sign, the current approach appears to be to improve machine learning by developing it to resemble human learning; to write an algorithm that edges ever closer to human cognition, namely, achieve the dream of 'strong' AI.

But perhaps that is neither the question to ask, nor the appropriate task at hand. What is ultimately still not completely explainable is how algorithms think. And while this appears to be a technical question – to open the 'black box' – there are also other ways of arriving at an understanding to answer that question. John Seely Brown's (2017) words come to mind: 'We must also be willing to constantly reframe our understanding of the world. We must regrind our conceptual lenses, and regrind them often.' (n.p.). One of these 're-ground' conceptual lenses, as this chapter has argued, is the issue of intelligibility, insofar as the task of intelligibility is to make the unknown known. In this chapter, I have argued for an approach to an alternative intelligibility via a conceptual approach of de-familiarisation, one that shifts the terms of understanding away from the human to those of an Other. In making this argument, I am aware I have scythed through whole swathes of literature, if only hoping to at least demonstrate the broad contours of the argument. It is also clear that much more work needs to be done to hone this approach into a systematic methodology of a robust conceptual framework of intelligibility applicable to algorithmic systems. But the first step, at least, is taken in attempting to frame an alternative question. For, per this section's opening quotation by Deleuze, the question of our intelligent machines in human society is not about the technology, but always of our future.

Notes

1 Randomness has also been long associated with creative work by humans, such as aleatory poetry or music by the Surrealists.

2 More fully known as the Dartmouth Summer Research Project on Artificial Intelligence.

[3] A pigment containing sensory protein which, for many seeing animals, including humans, is located in the retina of the eye that converts light into an electrical signal.

[4] This is a neologism coined by Vertov, referring approximately to 'the quality of the cinema-eye', as noted by the editor and translator of *Kino-eye* (Vertov 1984).

[5] It should be clarified that, contrary to the media's hyperbole, human players have indeed played a fifth line move before and to productive results, as part of the move's strategy would be to emphasise influence and speed to battle for more centre territory in return for giving up peripheral territory, akin to a chess gambit of giving up a pawn piece for centre control. The more precise reading here might be that the *context* for that strategy (for centre territory) was not present in this game, which was what made AlphaGo's move so strange and thus 'creative', rather than to claim that it was an 'inhuman' move.

[6] There are three networks to the algorithm: *the policy network*, which 'come up with what would be the interesting spots to play' to build up 'a tree of variations'; *the value network*, which 'tells how promising is the outcome of [each] particular variation', and finally *the tree search*, which would look through different variations and 'try to figure out what will happen in the future'. *AlphaGo – The Movie*, 47:15–47:50.

References

Alpern, A. 1995. Techniques for Algorithmic Composition of Music. Hampshire College, Fall. Retrieved from: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.9364&rep=rep1&type=pdf.

Ananny, M. and K. Crawford. 2018. Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. *New Media & Society*, 20(3), 973–989.

Appadurai, A. 2014. *The Social Life of Things: Commodities in Cultural Perspective*. Cambridge: Cambridge University Press.

Bailey, J. 1996. *After Thought: The Computer Challenge to Human Intelligence*. New York: Basic Books.

Boden, M. A. (Ed.). 1996. *Artificial Intelligence*. San Diego, CA: Academic Press.

Boden, M. A. 2004. *The Creative Mind: Myths and Mechanisms*. 2nd ed. New York; London: Routledge.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Braidotti, R. and M. Hlavajova. 2018. *Posthuman Glossary*. London; New York: Bloomsbury.

Branigan, E. 2006. *Projecting a Camera: Language-Games in Film Theory*. New York: Routledge.

Brill, J. 2015. Scalable Approaches to Transparency and Accountability in Decisionmaking Algorithms: Remarks at the NYU Conference on Algorithms and Accountability. Federal Trade Commission, 28 February. Retrieved from: https://www.ftc.gov/system/files/documents/public_statements/629681/150228nyualgorithms.pdf.

Brown, J. S. 2017. Sensemaking in our Post AlphaGo World. Stanford University mediaX Keynote, February.

Brown, W. 2009. Man Without a Movie Camera – Movies Without Men: Towards a Posthumanist Cinema? In: W. Buckland (Ed.), *Film Theory and Contemporary Hollywood Movies*, pp. 66–85. New York: Routledge.

Burkholder, J. P., Grout, D. J. and Palisca, C. V. 2014. *A History of Western Music*. (9th ed). W.W. Norton & Company: New York.

Bush, V. 1945. As We May Think. *The Atlantic Monthly*, 101–108.

Carr, N. 2010. *The Shallows: What the Internet is Doing to Our Brains*. New York: WW Norton.

Cellan-Jones, R. 2014. Stephen Hawking Warns Artificial Intelligence Could End Mankind. *BBC News*. 2 December. Retrieved from: https://www.bbc.co.uk/news/technology-30290540.

Cohn, G. 2018. AI Art at Christie's Sells for $432,500. *New York Times*, 25 October. Retrieved from: https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html.

Csikszentmihalyi, M. 1988. Motivation and Creativity: Toward a Synthesis of Structural and Energistic Approaches to Cognition. *New Ideas in Psychology*, 6(2), 159–176.

Csikszentmihalyi, M. 1990. *Flow: The Psychology of Optimal Experience*. New York: HarperCollins.

Davies, S. 2011. Still Building the Memex. *Communications of the ACM*, 54(2), 80–88.

Deleuze, G. [1985] 1997. *Cinema 2: The Time-Image*, trans. Hugh Tomlinson and Robert Galeta, Minneapolis, MN: University of Minnesota Press.

Diakopoulos, N. 2016. Accountability in Algorithmic Decision Making. *Communications of the ACM*, 59(2), 56–62.

Dreyfus, H. 1972. *What Computers Can't Do: The Limits of Artificial Intelligence*. New York: HarperCollins.

Dreyfus, H. 1979. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.

Economist, The 2020a. Technology Quarterly: Artificial Intelligence and Its Limits. 13 June. Retrieved from: https://shop.economist.com/products/technology-quarterly-artificial-intelligence-and-its-limits

Economist, The 2020b. What's the Worst That Could Happen? 25 June. Retrieved from: https://www.economist.com/briefing/2020/06/25/the-world-should-think-better-about-catastrophic-and-existential-risks.

Foer, J. S. 2016. Technology Is Diminishing Us. *The Guardian*, 3 December. Retrieved from: https://www.theguardian.com/books/2016/dec/03/jonathan-safran-foer-technology-diminishing-us.

Gardner, H. 1985. *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books.

Golumbia, D. 2009. *The Cultural Logic of Computation*. Cambridge, MA: Harvard University Press.

Grusin, R. (Ed.). 2015. *The Nonhuman Turn*. Minneapolis, MN: University of Minnesota Press.

Guyau, J-M. 1880. As Reproduced in Friedrich Kittler. 1999. *Gramophone, Film, Typewriter*, trans. Geoffrey Winthrop Young and Michael Wutz, pp. 30–33. Stanford, CA: Stanford University Press (first publ. in *Revue philosophique de la France et de l'étranger 5* (1880), 319–322).

Hayles, N. K. 1999. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago; London: The University of Chicago Press.

Hofstadter, D. 2000. *Gödel, Escher, Bach: An Eternal Golden Braid*, 20th Anniversary Edition. London: Penguin Books.

Ishiguro, K. 2005. *Never Let Me Go*. London: Faber and Faber.

Kaufman, J. C. and Baer, J. (Eds.). 2005. *Creativity Across Domains: Faces of the Muse*. Mahwah, NJ: Lawrence Erlbaum.

Kleist, H. [1810] 1972. On the Marionette Theatre. Trans. Thomas G. Neumiller, *The Drama Review: TDR*, 16(3), 22–26.

McCarthy, J., Minsky, M., Rochester, N. and Shannon, C. E. 1955. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. August. Retrieved from: http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf.

McLuhan, M. [1964] 2013. *Understanding Media: The Extensions of Man*. Berkeley, CA: Gingko Press.

Metz, C. 2016. How Google's AI Viewed the Move No Human Could Understand. *Wired.com*, 14 March. Retrieved from: https://www.wired.com/2016/03/googles-ai-viewed-move-no-human-understand

Miller, A. I. 1992. Scientific Creativity: A Comparative Study of Henri Poincaré and Albert Einstein. *Creativity Research Journal*, 5(4), 385–414.

Miller, A. I. 2000. *Insights of Genius: Imagery and Creativity in Science and Art*. Cambridge, MA: MIT Press.

Miller, A. I. 2019. *The Artist in the Machine: The World of AI-Powered Creativity*. Cambridge, MA: MIT Press.

Moffat, L. 2019. Putting Speculation and New Materialisms in Dialogue. *Palgrave Commun*, 5(11), n.p. Retrieved from: https://www.nature.com/articles/s41599-019-0219-8.

Moravec, H. 1998. When Will Computer Hardware Match the Human Brain?, *Journal of Evolution and Technology*, 1(1), n.p. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.136.7883.

Newell, A. and Simon, H. A. 1972. *Human Problem Solving*. London: Echo.
Parikka, J. 2015. *The Anthrobscene*. Minneapolis, MI: University of Minnesota Press.
Pasquale, F. 2015. The *Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.
Sautoy, M. 2019. *The Creativity Code: How AI Is Learning to Write, Paint and Think*. Cambridge, MA: Harvard University Press.
Sawyer, R. K. 2006. *Explaining Creativity: The Science of Human Innovation*. Oxford: Oxford University Press.
Searle, J. R. 1980. Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3), 417–457.
Searle, J. R. 1984. *Minds, Brains and Science*. Cambridge, MA: Harvard University Press.
Silver, D. et al. 2018. A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go Through Self-play. *Science*, 362, 1140–1144.
Sobchack, V. 1992. *The Address of the Eye: A Phenomenology of Film Experience*. Princeton, NJ: Princeton University Press.
Tegmark, M. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Alfred A. Knopf.
Turing, A. 1950. Computing Machinery and Intelligence. *Mind*, LIX(236), 433–460.
Veale, T. and Cardoso, F. A. (Eds.). 2019. *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems*. Cham: Springer.
Vertov, D. [1922], [1923] 1984. *Kino-Eye: The Writings of Dziga Vertov*. In Annette Michelson (Ed.), trans. Kevin O'Brien. Berkeley, CA: University of California Press.
Virilio, P. 1991. *The Lost Dimension*, trans. Daniel Moshenberg. New York: Semiotext(e).
Wald, G. 1950. Eye and Camera. *Scientific American*, 183(2), 32–41.
Yan, S. 2017. Google's AlphaGo A.I. beats world's number one in ancient game of Go. *CNBC*, 23 May. Retrieved from: https://www.cnbc.com/2017/05/23/googles-alphago-a-i-beats-worlds-number-one-in-ancient-game-of-go.html.
Zausner, T. 2007. *Artist and Audience: Everyday Creativity and Visual Art*. In R. Richards (Ed.), *Everyday Creativity and New Views of Human Nature: Psychological, Social, and Spiritual Perspectives*, pp. 75–89. Washington, DC: American Psychological Association.

Filmographic References

AlphaGo – The Movie. Video file, 1:30:27. YouTube. Posted by DeepMind, 13 March 2020. https://www.youtube.com/watch?v=WXuK6gekU1Y